FEBRUARY
1 9 5 3
VOLUME 17
NUMBER 1

JOURNAL OF CONSULTING PSYCHOLOGY

AMERICAN PSYCHOLOGICAL ASSOCIATION

JOURNAL OF CONSULTING PSYCHOLOGY

Edited by Laurance F. Shaffer

Teachers College, Columbia University

Associate Editors: Edgar A. Doll, *Devereux Schools, Devon, Pa.* William A. Hunt, *Northwestern University.* E. Lowell Kelly, *University of Michigan.* Morris Krugman, *Board of Education, New York, N. Y.* Bertha M. Luckey, *Board of Education, Cleveland, Ohio.* Ann Magaret, *University of Illinois.* Fred McKinney, *University of Missouri.* Anne Roe, *La Jara, N. M.* Carl R. Rogers, *University of Chicago.* R. Nevitt Sanford, *University of California.*

Lorraine Bouthilet, *Managing Editor*

The *Journal of Consulting Psychology* is the clinical journal of the American Psychological Association. It is devoted primarily to the publication of original research relevant to psychological diagnosis, psychotherapy and counseling, personality, and the dynamics of behavior. Case studies, theoretical contributions, descriptions of clinical techniques, and discussions of the professional problem of clinical psychology also appear in its pages.

Editorial Office

Manuscripts submitted for publication, books and tests for review, and correspondence on editorial matters should be sent to the Editor, Laurance F. Shaffer, 525 West 120th St., New York 27, N. Y.

Information for contributors: The important characteristics of a good article are the significance of its subject matter, the demonstrated soundness of its conclusions, its clarity, and its brevity. Clarity is improved by careful outlining, discriminating use of headings, and a simple, straightforward style of writing. For brevity, authors should strive to convey every thought without waste of words, and should avoid repetitions of methods, data or findings.

The style of all manuscripts should conform to the instructions given in the Publication Manual of the American Psychological Association, prepared by the Council of Editors (*Psychological Bulletin*, 1952, 49, 389-449). If it is not otherwise available, the *Manual* may be purchased, for $1.00, from the Washington office of the Association. The Editor expects each author to study the *Manual* carefully before submitting a manuscript. Special attention should be given to the style of headings, capitalization, indention, spacing, and footnotes, and to the preparation of tables, figures, and lists of references. Otherwise acceptable manuscripts which depart seriously from the standards of the *Manual* are returned to the author for revision and retyping.

There is a lag of about nine to twelve months between the acceptance of an article and its publication. Authors may secure immediate publication, however, by arranging for articles to be printed as extra pages, increasing the number of pages that subscribers receive. The cost to the author of such early publication is approximately $14.00 per printed page.

Authors receive fifty reprints gratis, and may order additional quantities when returning proof. Authors are charged one-half of the extra cost of tabular material and cuts, and the full cost of changes made on the proof.

Business Office

Communications concerning subscriptions, change of address, claims for the nonreceipt of a number, advertising, and other business matters, should be sent to the American Psychological Association, Inc., 1333 Sixteenth St., N. W., Washington 6, D. C.

The Journal of Consulting Psychology is published bimonthly in February, April, June, August, October, and December, the yearly volume comprising approximately 480 pages. The subscription per year is $7.00, foreign $7.50, single number $1.25.

Copyright, 1953 by the American Psychological Association, Inc. Printed and mailed by the Dentan Printing Co., 23 West Colorado Ave., Colorado Springs, Colo. and entered as second-class matter at the Post Office at Colorado Springs, Colo. Acceptance for mailing at special rate of postage provided for in section 538, Act of February 25, 1925, authorized May 27, 1940.

JOURNAL OF CONSULTING PSYCHOLOGY

February, 1953 Vol. 17, No. 1

CONTENTS

Pause Analysis as a Short Interviewing Technique: *Walter Toman* - - - - - - - - 1

Prediction of the Results of Therapy by Means of the Rorschach Test: *Lawrence S. Rogers and Kenneth R. Hammond* - - - - - - - - - - - - 8

Rorschach Criteria for Predicting Continuation in Individual Psychotherapy: *Benjamin Kotkov and Arnold Meadow* - - - - - - - - - - - - - - - - - - 16

Intelligence, Verbal Fluency, and the Rorschach Test: *Erwin J. Lotsof* - - - - - - - 21

The Effectiveness of Rorschach Elements in the Discrimination Between Neurotic and Ambulatory Schizophrenic Subjects: *Glenn W. Rieman* - - - - - - - - - 25

Productivity on the Rorschach as a Function of Order of Presentation: *Louis Joseph Maradie* - 32

Rorschach Characteristics of Persons Scoring High and Low in Acceptance of Self: *Robert E. Bills* - 36

Adjustment and the Discrepancy Between Self Concept and Inferred Self: *A. D. Calvin and Wayne H. Holtzman* - - - - - - - - - - - - - - - - - 39

Recent Research with Personality Inventories: *Albert Ellis* - - - - - - - - - - - 45

Distortion in Paranoid Schizophrenia: *Paul G. Daston, Gerald F. King, and Stewart G. Armitage* - 50

Some Effects of Sodium Amytal on "Deteriorated" Schizophrenics: *Willard A. Mainord* - 54

Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: An Exploratory Study: *Betsy Worth Estes* - - - - - - - - - - - - - - - 58

A Comparison of the Wechsler-Bellevue and ACE Tests on a University Counseling Center Group: *Reed M. Merrill and Louise B. Heathers* - - - - - - - - - - 63

The Effect of Variations in Ambiguity on Projection in the Children's Apperception Test: *Edith A. Weisskopf-Joelson and David B. Lynn* - - - - - - - - - - - - 67

Sex Sequence in the Draw-a-Person Test and Its Relation to the MMPI Masculinity-Femininity Scale: *Samuel Granick and Leon J. Smith* - - - - - - - - - - - - 71

A Note on the "Tree and Trauma" Interpretation in the HTP: *Murray Levine and Eugene H. Galanter* - - - - - - - - - - - - - - - - - - - 74

New Books and Tests - 76

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Pause Analysis as a Short Interviewing Technique

Walter Toman
Harvard University

The basic rule for the patient in psychoanalytic treatment is to say everything that comes to his mind, without selection [1, p. 23; 2, p. 192]. He is to associate freely, and whenever he stops, whenever he makes a pause, the psychoanalyst is, in principle, permitted to prod the patient. Thus the pauses the patient makes become specific points of interaction, and it seems quite possible that some of the peculiarity of psychoanalytic treatment when compared to any other form of psychotherapy springs from this procedure.

In prodding the patient to express what went on in his mind during the pause, or what idea interfered and made him stop, the analyst risks diverting the patient from the ongoing trend of free association, and at times it might be more important, therapeutically, to prod the patient to go on. At times the interruption by the analyst may be quite welcome to the patient, and the welcome topics are not always the ones psychoanalysis is after. In addition we should realize that it is when the patient pauses, that interpretations by the analyst or even questions leading up to interpretations are made. Seldom is such a question raised or an interpretation started by the therapist while the patient is still talking. Thus with a question referring to anything but the pause, we miss the pause, and with any question referring to the pause, we might miss further associations. It takes considerable skill to handle these intricate "molecular" interactions occurring in psychoanalysis, and this skill, in turn, requires training and experience, and experience and training, ad infinitum.

In 1950 H. Rohracher of the University of Vienna suggested an explanation of pauses in speech. While Johanna Karger started working on the quantitative and classificatory analysis of pauses in reading and speaking [3], I began to labor on the dynamic aspects. The reasoning given above was the starting point. Before I describe my proceedings, I would like to give a brief account of Karger's findings.

Twelve subjects had to read a short text aloud. They were sound-recorded and the pauses were explained afterwards by the subjects. Ninety-three per cent of the pauses were connected with punctuation marks, the rest with disturbances of task-awareness, distraction of attention, difficulties in apperception and conflicts between punctuation marks and personal organization of the text. The longest pauses varied between 0.5 and 1 second, and never exceeded 1.8 seconds. In free speech as contrasted with reading (instruction: "Explain why you chose your field of study or your particular profession.") the longest pauses exceeded 1.8 seconds with every subject. It was found that the faster the subject speaks, the greater the relative and the absolute average length of his or her pauses. The ratio of pauses made at all and pauses which the subjects recognized as pauses and explained was 6:1. The estimate of number of pauses made at all was based on the minimal time interval between two spoken words that a keen observer would just begin to consider a pause. This estimate depended on the rigor of the observer, and my ratio of 54.7:17 reported below does not contradict the former ratio. It only indicates a different standard as to what might possibly be considered a pause, and this standard is not crucial for my experiment. The causes of pauses in free speech were: (*a*) disturbances of composition, (*b*) distractions of attention, (*c*) disturbances in production of ideas, (*d*) pauses due to ordering the material, (*e*) time-consuming will processes, (*f*) disturbances in verbal expression (breathing, fatigue, slips of the tongue, etc.). No further specifications were attempted by J. Karger.

The Experiment

My experiment called for a life history of 5 to 10 minutes length, to be spoken by the subject into the microphone of a sound recorder. The subjects were given a casual half-minute of preparation, during which the experimenter got the sound recorder ready, or pretended to do so. Then the subjects spoke

their life history into the microphone. They were asked to take about 5 to 10 minutes, and to make it not too factual. The experimenter was present but did not interfere, not even by nodding or by any changes in facial expression.

After the life history was completed, the experimenter and the subject went over the record and analyzed the pauses. Only then did the subjects learn that the pauses were the object of investigation. About three-fourths of the subjects were surprised by the fact that apparently so much had gone on "between the lines" of their performance that otherwise would have escaped their notice.

Results

All of 35 subjects ranging from 18 to 34 years (13 male and 8 female students, 6 male and 3 female adults in nonacademic professions, and 3 male and 2 female adults in academic professions) revealed content material hidden or suppressed in the pauses that probably could not have been brought to light by other interviewing procedures in the given time, i. e., 5 to 10 minutes talking and 10 to 30 minutes of analyzing the pauses of the monologue. Subsequent to the test an interview was conducted with every subject in order to enable the experimenter to further appraise the significance of the content filled into the pauses by the subject while going over the record with the experimenter. A few examples of the performance will be given below.

The subjects produced life histories of an average length of 675 words, with a standard deviation of 230 words. They made an average of 54.4 pauses (with a standard deviation of 26), of which 17 pauses on the average (with a standard deviation of 9.2) could be interpreted by the subjects other than by mere formal difficulties of phrasing or searching for appropriate words.

In order to obtain some sort of an objective measure for the significance of the content material given by the subjects when interpreting their pauses, they were unexpectedly asked to come for another session 1 to 3 months later. Then they were exposed to their records again and asked to interpret the pauses again as far as they could still remember them. On the average 5.3 pauses were recalled (with a standard deviation of 2.7). The content of the other pauses had been forgotten. It turned out that 2.8 of the recalled pauses, i.e., of the recalled ideas or images or thoughts that had originally interfered with the delivery of their life history, were apparently affect-laden early memories, even very early childhood memories, that had oddly popped into the subjects' minds while speaking into the microphone. The rest were recent memories concerning the subjects' present activities and occupations, recent personal relationships and thoughts pertaining to the performance of the test.

These latter memories were presumably not too difficult to recall nor even to reconstruct from the context; but the earlier memories that would appear out of context of the delivered life history to anyone but the subject himself or a sophisticated psychologist would have been impossible to recall unless the same dynamic factor that had originally brought them up had been reactivated. These memories seem to stand for stronger (older and possibly unconscious) dynamic constellations. We may conclude that they were significant, important memories, otherwise the subject would not have recalled them on the right spot in the objectively inconspicuous context of his delivered life history. In that sense this check may be considered an objective measure for the significance of the content material given by the subjects when interpreting their pauses.

Four examples of records interpreted immediately after the test are given below. All further references and comments to the cases have been omitted. They will serve only as a demonstration of the test performance including the interpreted pauses. Thoughts going on during the pauses and reported by the subjects while going over the sound record with the experimenter are underlined.

Examples

Subject A (male, 20 years old)

I remember very little of the first few years. Just that I was born in K. and moved immediately to D. where I lived until I was about 5 years old. I had a very . . . *a little scene in kindergarden. A little girl was crying lustily when having to stay there. I cried too. We made quite a scene . . .* strong, childish attraction for a little blonde girl there, and we have kept writing ever since, and

she came to K. last, last month, which was the first date I have had for, during my college experience. This particularly bothers me. Anyway, then we moved to E. I went to the local grammar school, very . . . *out in the country, worst teachers, poor reputation. Father worked to set up schools. Had many enemies. Mother has practically gone blind (over all this). She is very introverted and sensitive* . . . poor grammar school, actually, and there I had three good boy friends, and one pretty blonde girl again with whom I spent a number . . . *her older brother got between us. Had a tent. Wanted his sister and me to undress and play with each other. I never liked the fellow* . . . of the first few years. I can remember that I used to be terrifically sensitive in grammar school, particularly about playing sports. I was very slow on my feet. I could not run in the right way. I used to run on my toes instead of on my feet. So the others would goad me into playing tag with them, practically every day during the first 6 years of grammar school, it seems, anyway. And I would get . . . *I got very hurt and nervous. My best friends were very callous. Were best friends, and stabbed you in the next minute. I was worried about other people's opinion about me.* . . . I tried not to show my anger and, but it did affect me very much. I was very easily offended and hurt. I remember one time that one boy who was actually my best friend made a nasty remark, and I had to stay home that afternoon, because I could not. . . . *I lowered myself by offering him a bicycle ride home. Felt disgusted with myself* . . . face going back to school again. Then going to high school I did very well in school. I was number one student in town graduating to high school and so I had a very easy time. My brother went to high school the year before, and he made such a terrifically good impression that I got by on his own scholastic ability, that is, they gave me all A's without really knowing. . . . *I felt inferior to my brother. But then I became the first in town. My brother had left town* . . . what my own talents were. Although I don't know whether I would have gotten A's or not. Probably yes. But I never had a date with a girl during high school and this particularly bothered me. I never learned to dance, and I felt that was one of the most unfortunate parts of it. . . . *My parents should have encouraged me. They let me down. Brought me up in the wrong way. Later on I talked it over with my father. He said: "We never stopped you." But I had the impression, they did not approve.* . . . My parents during this time never encouraged me or inhibited me from going with girls, but I never had the courage to do it myself, and it has carried over until now, practically. I suppose this is the most interesting part of what I am supposed to say. I don't know. And then I wanted to get on the football team in high school, but never did actually try out for it. Perhaps I could have made it. Then during my junior year of high school my parents decided they wanted me to go to A. for my senior year, and so I had to work very, very hard then in order to make the scholastic standing necessary to get into A. I did get into A. with full scholarship and felt pretty good about my own capabilities then. But when I got there I found that in the first term in A. I got a C minus-and-a-half average, which was very, very poor, particularly for a scholarship boy, so after that I became a recluse and spent about all of my time studying . . . *there were dances in A., Saturday nights. I wanted to go, but did not. Used to sit in my room* . . . so that I had no social life to speak of at A. I never joined a club or anything. I just stayed in my room and ground away. Then I . . . *My brother was in college. I was just going to follow him. I was not very ambitious to go. I just as soon would have worked* . . . did not give much thought to college. I did not really expect I would be able to go to college, because of financial reasons, but I half-heartedly applied, took the college entrance exams, and U. admitted me. It was the only college I had applied to. And . . . *they should have (awarded me a scholarship). I had good records, eventually. A. did not elect me for the cum laude society. They slighted me. They sent me to Washington, free, for a week, though. My brother had been elected.* . . they did not award me a scholarship, though. The only way I came here was that A. managed, A. did give me a prize scholarship, $400, and also the X.Y. Association gave me a scholarship. And also a very old lady who just died gave me a gift of $2500 for my school, my college career, which I could not have done without, and I did finally get to U. . . . *money stretched over 4 years made things easier, but still pretty hard.* . . . And during my freshman year in U. I roomed with a Jewish boy who was about as introverted as I was, and I found myself not particularly liking him, but pitying him in a way. . . . *in French class two girls were sitting next to us, but we never approached them* . . . and we became rather good friends. It was a rather artificial friendship, since I . . . *this amounts to saying, I am superior. His father (who had invented something and discussed advertising it with his son) said: "We put a nice nude on it." The boy had some nasty habits, was dirty, always picking his nose* . . . pitying him, as I did, it was a rather superior feeling I had towards him. Nevertheless we did become very good friends. We, we both wanted to go out with V. girls, but we never managed it. I managed to rationalize myself into thinking that it was because of my low finances, and that's partly true, I suppose . . . *it was in me, of course* . . . but the result was what I figured a rather poor sexual adjustment. And being a senior now this has carried over. And I have had one date in my life time, so to speak, that is a couple during . . . *my and her parents were along. She was older than I was* . . . during the summer

times, picnics and that sort of things, but I find it very, very difficult to approach girls in a. . . . *I am getting all worked up. This problem is a contemporary one* . . . in a way that's not self-conscious.

Subject B (female, 19 years old)

I was born in N., and I can't remember anything until I was about 3 years old . . . *memory of the house we used to live in: picture of my uncle standing in the toilet; father said: "He is taking a picture. Come on!"* . . . and I can remember being a happy child. I think I got a lot of attention. . . . *I was over-protected, I guess* . . . from my parents when I was very young. And I had a lot of success in school, and even in kindergarden I remember I was the best reader . . . *from when on was I a lucky child? What facts do I have?* . . . and I could draw and paint . . . *this is not enough. What did I do at home? Can't find anything* . . . and all that sort of thing. And I went to elementary school. In about the third grade, I suppose, I was, oh, 9 years old, I had a little trouble with the children in my class. I was much more withdrawn. I was more imaginative, introverted than they were. I would not really understand the more social gang-like times that they had. . . . *I felt left out. I used to cry. Mother was so helpful. She handled me so good* . . . and I felt left out. I felt very unhappy. I remember they used to make fun of me. And I was always the kind of kid that used to go by herself and, oh . . . *fairies, later on love, also nature, lose myself in those things. Reading and writing poems. Thoughts like: How do you feel the infinite? What happens if you fall forever?* . . . imagine things a lot. Oh, I wrote poetry starting from when I was 6 years old, I think. And I went to . . . *much earlier than high school I became a social being. I left out girls, got along very well with boys, was never in a gang. The girls used lipstick, they were teen-agy quite early* . . . a very, very fine high school in N. I always had a lot of academic success, and in spite of this early withdrawnness of mine, I guess, I was fairly popular, and in high school I majored in art, and I became interested in dancing and music and all these other things, and I had a wonderful time, a lot of social success, academic success. I became good in many, many things, and I really, really enjoyed school aaa . . . *so much to tell. People were so congenial, liberated, very alive and lively atmosphere, good relationships with teachers* . . . a great deal. As far as my home life went at that time, I think, starting around, oh, 15, I began. . . . *I wanted to probe everything. Not deny any experience. Look for the truth.* . . . I began to rebel quite strongly against my family, because I did not think that what they demanded from me was what I wanted to fulfill. And I feel now, that they were very, very . . . *there was a fight. I did a lot of square dancing. Wanted to go with a Negro boy. They forbid it, but I did it. There was hell in the house* . . . oh, they expected a lot in the way of nominal achievement and following what their goals were, I think, much too much. And starting at 16 I began to rebel. I became critical of American culture. I think, yes, and this was also the same time when I thought I was discovering that I was popular and that I had a nice personality, and that the boys liked me, and that sort of thing. And. . . *nowadays I dislike my adolescence. So much dishonesty in that age, effort to make impression, feminine wiles. Disgusting* . . . this went to my head for a little while, and after that I became, as I said, critical and went to the other extreme. I joined a, aaa . . . *how should I characterize it, express the spirit?* . . . youth Zionist group which was very, very, very revolutionary in its mores and, oh, no lipstick and girls treated completely equal with men. . . . *These were the goals, but they did not really achieve them. Girls became masculine. No positive culture of their own* . . . oh, really a little community within itself, and Marxist orientation . . . *they were not even good Marxists* . . . and tried very hard to establish honesty and introspection, and complete naturalness . . . *some of the girls opposed integration, thought Zionism is the solution, quite doctrinaire* . . . I guess, as goals. And oh, when I came to C. as a freshman, I was just, just coming out of this. . . . *Girls in C. less accepting than in N., pressure toward conformity; it was hard to accept them as people. I had a real war for a year.* . . . I was always somewhat critical, and yet I found things there I felt, and still feel, . . . *competition, selfishness, sterility of the whole culture* . . . are lacking in American culture. And I became very interested in Judaism as a way of life, and politically a Zionist, and, well. . . . *Oh, I am up to the present* . . . at this point I no longer am a political Zionist. I am very interested in Judaism. I am interested in modern dancing. I do more of that, practically, than anything else. I have made a conscious effort to narrow down my . . . *talents? That's conceited* . . . fields of expression and chosen dancing as the one I more or less concentrate on. I hope after college. . . . *well, I don't really know. It's all vague* . . . somehow, oh yes, I have always been interested in. . . . *I left out my settlement experiences. Do I really have abilities in therapy?* . . . therapy and helping people. I did a lot of work in settlement camps during the summers with a group of kids. I like to teach very much. I enjoy working with children, and I would like to combine therapy with dancing. And possibly go to Israel after college. I guess that's about all.

Subject C (male, 27 years old)

. . . *Shall I tell him my mother was psychotic, or not?* . . . . Well, I think, one of the things that's important in my life history is the. . . . *I am going to tell him that, but I start somewhere else* . . . is the, oh, type of place of my childhood. I grew up

in a very small town in the West, in a culture quite different from the one that I am in now. . . . *I felt better in childhood than now. Is that neurotic?* . . . aaa, life there is quite in contrast to the one life one leads around here. It's a rather quiet, free and easy sort of existence, whereas everything here seems to me to be very hostile, full of pressure, aaa. . . . *What shall I say next?* . . . to accomplish things all the time. I find considerable lack of time to sit back and relax and sort of give in. . . . *I better not get emotional* . . . to the enjoyment of . . . *existence? living? that's a poor sentence* . . . my existence. Aaa, another very important thing in my life are my parents, aaa. . . . *Whom shall I take first?* . . . . My father was a very quiet . . . *image of him* . . . retiring sort of person, very, very conscientious, working extremely hard, and not very . . . (*some thought that subject could not recall*) . . . had not very much to say when he was at home. Seems to me, he was a sort of a shadowy background figure in the household rather . . . *but he was warm, even though reticent* . . . than a dominant person. My mother . . . *her father very stern, militaristic, had a lot of control, her affective needs starved* . . . grew up on a farm in the West, had, I think, a rather hard childhood herself that carried over to her later life, and . . . *unfortunately, although I should not say so* . . . to my life. I think, she tried. . . . *I better try to understand* . . . overly hard to compensate for some of the things that she missed in her childhood, and as the result of that she tried too hard to make things easy for myself and for my brother . . . *image of myself as a child of 6 or 8* . . . tried too hard to . . . *control* . . . protect us from the vicissitudes of life, rather had a certain, had a negative effect on my life. I think I lack a lot of . . . *image of a very energetic, goal-directed person* (*male*) . . . drive and autonomy that I might otherwise have had . . . *some sort of therapy might help, I hope* . . . aaa. I think, probably the most significant thing in my life that I can think about, if I were to point to a single incidence, was the psychotic break that my mother had when she was about etc.

Subject D (female, 21 years old)

The first thing I remember is wanting to go to school and starting a little early, and liking it very much . . . *picture of myself, over-confident, boastful* . . . and all during grammar school I was pretty sure of myself and liked it. And when the first . . . *my father did not want me to go to school. My sister went. It was too hard. I was not strong enough, they said* . . . conflict came, when it became time to choose a high school; I chose one that was strictly college preparatory, to which girls from all over R. went, as opposed to a local high school. And as soon as I started that I lost confidence . . . *picture of myself. What would it be to others, when I make new friends?* . . . felt more pressure upon me. I had to work harder, was not sure of myself as far as making friends went, and although I liked school very much, I disliked the teachers, the pressure I felt from them. . . . (*After 10 seconds*) *I hate not to be able to say anything* (*further 20 seconds elapse*). . . . It was an all girls' school which was nice from the point of view of having girl friends, though we were always quite shy about meeting with boys and did not like to go for that reason. And the school put a lot of pressure on it, towards choosing a good college, towards making us want to go, but never gave us a chance to decide whether or not we wanted to go. It just assumed that we would. And upon going to college I have never again felt as much pressure as I did in high school and was glad to go to a college that is at least partially coed. I also felt that the situation was a lot more natural.

Discussion

There seem to be certain advantages to this technique. One of them is the following: The subjects tend to organize their thoughts as well as possible and to give a somewhat coherent story. The fact that their performance is sound-recorded increases their concern with the quality of their performance. On the other hand, they are not given sufficient time to organize their story in advance. Rather they are caught by the sound recorder in the process of organizing, and in this sense the technique may be called a situational test. All criteria developed for appraising verbal creations are eligible and may be applied.

The effort of organizing the story for the sound recorder is presumably greater than the effort of organizing it for a listener in a more familiar situation like interviewing, or counseling, or any conversation. A little speculation may be inserted here, that has been confirmed on other grounds by the author. If we assume that a higher effort put into the performance would consume more "psychic energy," we would expect less energy to be available for any concurrent activities. The main concurrent activity in the delivery of one's life history consists of the defenses against making slips, against revealing too much or uncovering one's emotions more than intended. We would predict that the greater effort may either lower the effectiveness of the defenses against unwanted impulses or, if this is not the case, the defense mechanisms themselves may become more apparent. Both types of effects may give the counselor or therapist cues about the sub-

ject that he might not be able to secure otherwise, at least not in the given time or not simultaneously with a life history of the subject. Examples of the first type have been given (subjects A, B, C). Examples of the second type are a few meagre life histories among my records, with long pauses and frequent blanks with respect to the pauses (subject D). These subjects reported that realizing that they made pauses extended one or the other pause to an unusual length, and this in turn blocked their performance, and caused further pauses, until they eventually gave up.

Another advantage becomes clear when we compare this technique with psychoanalytic therapy in particular, but also with any counseling or interviewing technique. As pointed out in the beginning, the pauses are crucial events in psychoanalytic treatment, and however the therapist deals with a single pause, he is interfering. The reported technique yields a record of the uninterrupted cognitive processes of 5 to 10 minutes in length, a sufficient time for a fruitful systematic analysis of the meaning of the pauses.

A third advantage becomes apparent when the record is reviewed. The subjects are surprised by what they said, even surprised by their voice, if they have not heard it recorded before, and this surprise contributes strongly to their willingness to cooperate in the analysis of their pauses. They feel encouraged to take a more objective attitude toward what they said into the microphone than they do when reference is made to what they said in an ordinary interview or conversation with no recording. All subjects except the few who had given very meagre life histories with extremely long pauses were willing to talk very frankly and openly about themselves and their lives, but even the defensive subjects were willing to talk at least about their inhibitions and defenses and found some relief in so doing. Subject and experimenter (client and therapist) are both dealing with the record of the subject's past performance, and this may account for the objective attitude the subject is able to assume quite easily. According to an accepted psychoanalytic metaphor one part of the ego has to cooperate with the analyst in getting a hold on the unconscious parts of the ego (and on the other systems). The pause-analysis technique facilitates this cooperation. The subjects slide into a good contact and rapport with the counselor or therapist, while talking about their performance, their pauses, and even their defenses.

A fourth advantage is the possibility of going over the subject's performance at any later time. What pauses are recalled that have been interpreted (given content and meaning) by the subject in the first session may give additional cues (see "Results" above). Furthermore, the attitude of the subject to his now quite remote performance can be appraised and utilized in several clinical aspects.

Do the subjects manage to interpret all the pauses they make? I would say, they manage to interpret all the pauses that are important as to contents and dynamics. Of 54.4 pauses on the average counted by the experimenter, 17 pauses on the average were commented on and interpreted by the subjects. All the other pauses could safely be considered as pauses due to formal difficulties, difficulties of finding the appropriate words or of finishing a sentence. The subjects were asked to tell what went on during the pauses while hearing their records, and in the beginning the experimenter stopped the record at whatever he considered a pause, but soon the subjects began to cooperate quite efficiently by telling when to stop the record and when not to stop it. Among other reasons, this could be relied on since the subjects had been told that they could refuse to talk about things they did not want to confide. In this case they had only to indicate that there were such things. However, very few subjects resorted to that privilege.

Can we believe what the subject reports? This question may be raised just as well with any psychological approach to personality, and particularly with any form of interviewing. On the grounds mentioned above (good rapport with experimenter, increased objective interest, etc.) we may expect more reliable reports from the subject with the pause-analysis technique than with other short interviewing techniques.

Why apply the analysis of pauses to life histories and not, say, to TAT stories? There is no reason why one should not also apply it successfully to TAT stories. But the life

history of a subject is probably the most crucial and important of all the stories he can tell. Any therapeutic effect can, in some sense, be viewed as a recentering, a revision of the life history of the subject, and it is not the events that are revised, but their evaluation. The disappearance of neurotic symptoms is only a secondary effect. Life histories, verbal or written, given by the patients before and after treatment may in general be the most valuable and easy-to-obtain, although certainly imperfect objective checks on these revised evaluations, on a changed self concept, on a liberated and strengthened ego, or on whatever else a therapeutic success may be said to be based. We may also maintain that life histories can be produced repeatedly with interferences by previous performances of no greater intricacy than we would find with other complex clinical tests.

All the shortcomings of this method, such as reliance on what the subject himself tells about the pauses, too many variables, too little quantification, no complete elimination of social interaction and of its interference with the performance to uncontrolled degrees, etc. hold, in one way or another, for any clinical test. In the pause-analysis technique the emphasis is on contents and dynamics and their clinical appraisal. Tendencies pertaining to the person's (subject's, patient's) specific structure and organization of the mind, but also to his performance and to the test situation can be traced via contents of the pauses and its placement in the context of the delivered life history, and any information that could not have been secured by other techniques in the same length of time will be valuable in clinical practice. The technique seems particularly promising in cases on which relatively little time can be spent.

Summary

A clinical technique has been described in which the subject is asked to tell his life history into the microphone for 5 to 10 minutes. After that the subject and the experimenter (counselor, therapist) go over the record and analyze the pauses the subject made. Some quantitative results of 35 subjects have been mentioned and four examples given. Theoretical implications, advantages and disadvantages of this technique have been discussed.

Received June 23, 1952.

References

1. Fenichel. O. *The psychoanalytic theory of neurosis.* New York: Norton, 1945.
2. Freud, S. The interpretation of dreams. In *The basic writings of Sigmund Freud.* New York: Modern Library, 1938.
3. Karger, Johanna. Psychologische Ursachen von Sprechpausen. Unpublished doctor's dissertation, Univer. of Vienna, 1952.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Prediction of the Results of Therapy by Means of the Rorschach Test[1]

Lawrence S. Rogers
Veterans Administration Mental Hygiene Unit, Denver, Colorado and University of Colorado

and Kenneth R. Hammond[2]
University of Colorado

One important problem facing mental hygiene clinics is attempting to evaluate before beginning therapy whether or not the patients have the capacity to profit from psychotherapy. Much time and energy ordinarily devoted to the unsuccessful treatment of cases might be more profitably diverted to the better therapy risks should it be possible to make such an evaluation correctly. Our problem, therefore, was to investigate the possibility of using the Rorschach test for prognostic purposes. In short, can the Rorschach test be used to predict which cases will be successful in therapy in a mental hygiene clinic?

Method

The method used involved establishing criterion groups: one group composed of patients who were successful in psychotherapy; the other group composed of patients who were unsuccessful in psychotherapy. The first step was to proceed *deductively* to differentiate between Rorschach protocols of patients in the criterion groups. The next step was to proceed *inductively* to discover whatever differences might exist between the criterion groups and then test these differences on a new sample.

[1] Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are the results of their own study and do not necessarily reflect the opinion or policy of the Veterans Administration. This paper was read at the 1951 meeting of the American Psychological Association.

[2] Also contributing to this research were Lewis Bernstein, David Ficks, Lillian Sandberg, and Joanne Knauss.

Our deductive procedures included first, a "total impression" judgment. That is, psychologists read the patient's Rorschach protocol and predicted whether the patient was a member of the successful or unsuccessful group. Secondly, "signs" derived by Harris and Christianson [2] were utilized to assign the Rorschach protocol to either criterion group. Thirdly, predictions toward the criterion were made on the basis of whether the *M* present in the protocol was of the flexor or extensor type. Fourthly, two different combinations of Rorschach variables thought likely to discriminate between the criterion groups were tried. The above four procedures, then, were a priori.

Our inductive procedures followed the usual pattern of ascertaining differences between "known groups." First, various Rorschach scores or variables were compared in the two groups. This comparison was made in terms of the percentage of persons in the criterion groups having a given score on a given variable. Secondly, patterns were developed from an analysis of variables which seemed promising as discriminators and these patterns were applied to a new sample. The above two procedures, then, were a posteriori.

Procedures and Results

The procedure began with the establishment of the two criterion groups. One group consisted of 50 protocols of patients whose case file showed that at termination they were thought "unimproved." The other group consisted of 59 protocols of patients terminated "slightly improved" or "improved." The pro-

tocols were rescored according to Klopfer's method — using, however, Beck's tables of $F+$ and popular responses. All cases were selected at random from the clinic files, without regard to diagnosis or therapist. All subjects in the criterion groups, as well as in the cross-validation groups, were veterans who had had treatment at a Veterans Administration Mental Hygiene Clinic.

The judgments concerning the success of the patients in psychotherapy were made by the patients' therapist in consultation with a senior psychiatrist. The senior psychiatrist was one who had been regularly consulted on the case. There were eight social workers, eight psychiatrists, and three psychologists involved.

Deductive Procedures.

The deductive procedures were as follows:

Total impression judgments. Three psychologists, with Rorschach experience ranging from three to five years, judged each Rorschach protocol for success or nonsuccess under four different conditions — each condition separated by at least a one-month interval.

The procedure under Condition I was to make a *snap* judgment, in not more than two or three minutes, concerning success. The procedure under Condition II was to make a *studied* judgment, with as much time allowed as was felt necessary by the psychologist making the judgment. Condition III was the same as Condition II except that the name of the patient's therapist was given to the judge as additional data. Condition IV provided a check for Condition III in that the therapist's name was given to the judge without the Rorschach protocol. That is, predictions of success were made from the therapist's name alone.

Validity coefficients were calculated for each psychologist under each condition and were expressed in terms of chi square computed for a four-fold table.[3] All results from the total impression judgments were negative. Predictions were no better than chance under all four conditions. Results are presented in Table 1.

Table 1

Validity of Total Impression Judgments

Psychologist		Type of Judgment		
		Snap	Studied	Therapist
A				
	χ^2	.151	.054	.400
	p	>.60	>.80	>.50
B				
	χ^2	1.166	.156	.516
	p	>.25	>.60	>.40
C				
	χ^2	1.37	1.22	.035
	p	>.20	>.20	>.80

Interpsychologist and intrapsychologist agreement under each condition was also calculated and was expressed in terms of tetra-

Table 2

Tetrachoric r's Indicating Inter- and Intrapsychologist Agreement on Total Impression Judgments

(Italicized numbers refer to intrapsychologist agreement)

	Psychologist								
	A			B			C		
	Snap	Studied	Know thpst.	Snap	Studied	Know thpst.	Snap	Studied	Know thpst.
A									
Snap		*.75*	*.74*	.44			.30		
Studied			*.87*		.50			.21	
Know thpst.						.60			.32
B									
Snap					*.65*	*.65*	.51		
Studied						*.53*		.46	
Know thpst.									.27
C									
Snap								*.85*	*.80*
Studied									*.85*
Know thpst.									

[3] An objection to the use of chi square might be raised here on grounds that the judgment of one case was not independent of the judgment of another. An attempt to meet this objection was made by not telling the psychologist how many cases were in each criterion group and warning him that he was *not* to assume that the criterion groups were nearly equal in number. Thus, it was hoped that the psychologist judged each new case independently of how he had judged previous ones.

choric *r*. Tetrachoric *r*'s[4] for the judgments between different psychologists ranged from +.21 to +.60 and are presented in Table 2. Correlations above +.40 are significant at the 5 per cent level. Five of the correlations range above +.40 and they are to be considered significant. According to Wilkinson's table [3] the probability of obtaining five correlations significant at the 5 per cent level by chance when nine are calculated is less than .0006.

Tetrachoric *r*'s for the same psychologist under Conditions I — III ranged from +.53 to +.87 and are presented in Table 2. All of these correlations are significant at the 5 per cent level, although how contaminated they are from judgment to judgment is unknown. The judges expressed the not unlikely point of view that contamination from condition to condition was minimal because of the large number of cases (109) and the time interval.

Harris and Christianson's scoring method. Weights for various scores as derived by Harris and Christianson [2] in their study on the use of the Rorschach in the prognosis of therapy were applied to our data. This procedure did not discriminate between the criterion groups ($\chi^2 = .110$, $p = .75$).

M responses. An analysis was made of all the *M* responses with regard to extensor or flexor quality. It should be noted that the examiners did not press the subjects for the direction of movement; any such expression was completely unsolicited. *M*'s appeared in 76 records; in 30 cases there were only flexor *M*'s, in 13 cases only extensor, in 12 cases there were both, and in 21 cases the direction of movement could not be judged. Of the 30 cases in which there were only flexor *M*'s, 12 were improved cases, and 18 unimproved; of the 13 cases in which only extensor *M*'s appeared, 11 were improved and two unimproved. The presence of extensor *M*'s only, therefore, appears to be associated with improvement in therapy. These data seemed promising and are elaborated below.

Response patterns. An attempt to differentiate between the criterion groups on the basis of different patterns of responses was made. Four variables were chosen, not because in combination they represented a particular personality structure, but because of their importance as Rorschach variables. These variables were *R*, *M*, *F%*, and ΣC. A second procedure of this sort was made with five variables, *R*, *M*, *F%*, number of rejections, and presence or absence of additional responses. The second procedure differs from the first in that the last two variables mentioned were included because they were regarded as useful prognostic signs. The hypothesis was that those patients with fewer rejections and a greater number of additionals would be the better therapeutic risks.

A tabulation method was used to analyze the data. Space does not permit complete presentation of this method here. See Hammond [1] for a detailed description of the method.

In the first procedure each variable was broken into categories of High, Medium, and Low.[5] For example, Low $R = 0 - 15$, Medium $R = 16 - 26$, High $R = 26+$. For Low and Medium categories of *R*, *M* was categorized as follows: Low $M = 0$, Medium $M = 1$, High $M = 2+$. For the category of High *R*, Low $M = 2$ or less, Medium $M = 3 - 5$. High $M = 6+$. *F%* was categorized as follows: Low $F\% = 0 - 40\%$, Medium *F%* as 41 – 60%, and High *F%* as 61%+. ΣC was categorized in this fashion: For Low ranges of *R*, Low $\Sigma C = 0$, Medium $\Sigma C = .5 - 1.4$, High $\Sigma C = 1.5+$. For Medium ranges of *R*, Low $\Sigma C = 0 - 1.0$, Medium $\Sigma C = 1.5 - 2.4$, High $\Sigma C = 2.5 +$. For High ranges of *R*, Low $\Sigma C = 0 - 1.9$, Medium $\Sigma C = 2.0 - 3.4$, High $\Sigma C = 3.5+$.

For the second procedure, *R*, *M*, and *F%* remained as described above. Rejections were categorized in the same way for all values of *R*, Low rejections = 0, Medium rejections = 1, High rejections = 2+. "Additionals" were categorized merely in terms of presence or absence. Results from the pattern analysis procedure were negative. By inspection it was

[4] Splits closely approximate .50 — .50. Condition IV is omitted.

[5] Decisions about the range in categories were made from an examination of the data obtained. Although recognized as a risky procedure, it appeared to be the most promising one and risks would be forestalled by use of a cross-validation group, if positive results were obtained.

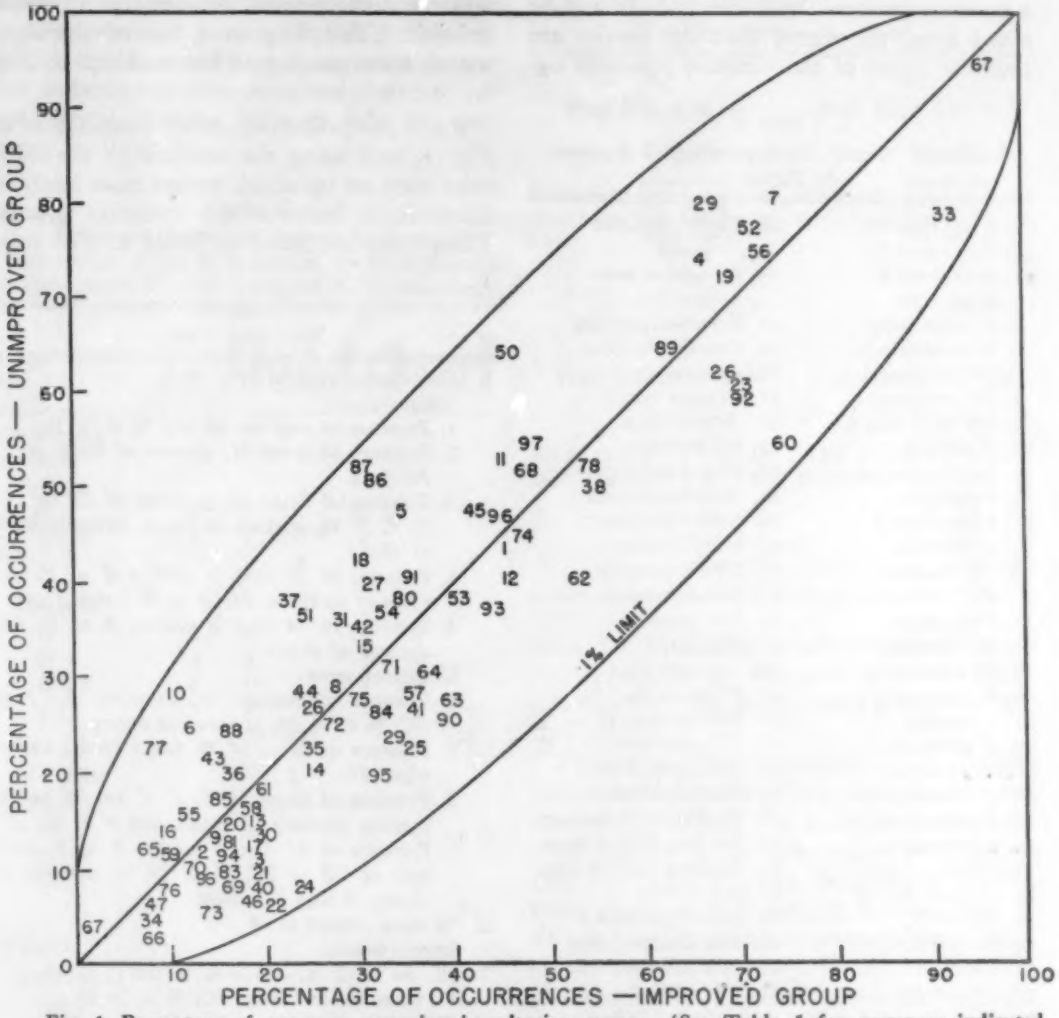

Fig. 1. Percentage of responses occurring in criterion groups. (See Table 5 for responses indicated by numbers in this figure, e.g., $1 = R < 20$.)

obvious that the patterns used did not discriminate between the criterion groups.

In summary, then, only one a priori method of prediction, type of *M*, seemed even promising; the other methods offered no evidence of better-than-chance success.

Inductive Procedures

The inductive procedures were as follows:

Single variables: The first step was the analysis of single variables, or scores, such as *M*, *ΣC*, *F%*, etc. Each such variable was examined to discover whether the percentage of persons achieving a certain score differed in the criterion groups. For example, the percentage of persons in the criterion groups who achieved three or more *M*'s was compared. The procedure for choosing cutting points was deliberately that of choosing those cutting points which seemed empirically most promising. There were two reasons for this procedure: (*a*) The study had now become a frankly empirical one and cross-validation was planned, and (*b*) it very soon became apparent that few large differences were going to be found.

Ninety-nine percentages were compared and are presented pictorially in Fig. 1 by means of

a method developed by Zubin [4]. It will be noted from the figure that the results are negative. None of the variables produced significant differences between the criterion groups.[6] Table 3 gives a list of the signs tested, corresponding to the numbers in Fig. 1.

Table 3

Rorschach "Signs" Corresponding to Numbers In Figure 1

1. *R* less than 20	52. Categories more than 6
2. *R* less than 10	53. Categories more than 8
3. *R* 40 or more	54. Rejection presence
4. *M* presence	55. Cut-off 3 or more
5. *M* more than 1	56. *At* presence
6. *M* more than 3	57. *At* more than 2
7. *FM* presence	58. *At* more than 3
8. *FM* 4 or more	59. *Sex* presence
9. *FM* more than 4	60. *N* or *Geog* presence
10. *m* presence	61. *Art*/*Des* presence
11. Add. *m* presence	62. Nature presence
12. *k* presence	63. *Geog* presence
13. *k* more than 1	64. Clouds presence
14. *K* presence	65. Blood presence
15. *FK* presence	66. Fire presence
16. *FK* 3 or more	67. $AD > A$
17. *F* 4 or more	68. *F*% more than 70
18. *F* 3 or more	69. *F*% more than 50
19. *Fc* presence	70. *F*% less than 30
20. *Fc* 3 or more	71. *F*% 60 or more
21. *c* presence	72. *F*% less than 40
22. *C'* presence	73. Shading absence
23. *FC* presence	74. Shading 15% or more
24. *FC* more than 2	75. Shading 20% or more
25. *FC* more than 1	76. Shading 30% or more
26. *CF* presence	77. *Sum C* = 0
27. *CF* more than 1	78. *Sum C* less than 2
28. *CF* more than 2	79. *Sum C* more than 4
29. $FC > CF$	80. *Sum C* less than 1.5
30. *FC* to 0 *CF*	81. $CF > FC$ & *Sum C* > 3
31. $CF > FC$	82. *M*:*Sum C*. Same as *OM*/*OC*
32. *CF* to 0 *FC*	83. % last 3 :24 or less
33. Presence of *CF* or *FC*	84. % last 3 :30 or less
34. Presence of *C*	85. % last 3 :45 or more
35. Human absence	86. % last 3 :40 or more
36. Human more than 3	87. % *C* less than 10
37. Human more than 2	88. % *C* 20 or more
38. *Hd* presence	89. % *C* 13 or less
39. *Hd* more than 1	90. *W*% > 50
40. *Hd* more than 3	91. *W*% 30 or less
41. $Hd > H$	92. *W*% < 20
42. *A*% 50 or more	93. *W*% 45 or more
43. *A*% 60 or more	94. *D*% 30 or less
44. *A*% 30 or less	95. *D*% 40 or less
45. Populars 5 or more	96. *D*% 65 or more
46. Populars 2 or less	97. *D*% 41 – 60
47. Populars 8 or more	98. *S* presence
48. $A > Ad$	99. *S* more than 1
49. *Aobj* presence	
50. *Obj* presence	
51. Others presence	

Rules for prediction. After examination of Fig. 1, and using the analysis of the *M*'s, rules were set up which seemed most likely to discriminate between the criterion groups. These rules are listed in Table 4. The rules were then applied to a new sample of 104 cases, of whom 70 were improved, and 34 were not. Upon application, it was noted that the rules were not mutually exclusive. It was, therefore, decided that each case would be judged only once. Since it appeared from our original sample that we could predict success more accurately than failure, if a case fell into a category predicting success, that category was the only one used. If a case fell into more than one category, it was put in the one with

Table 4

Prediction Rules

I. Considering type of *M*

Improvement:

1. Presence of extensor *M* and % $C > 10$
2. Presence of flexor *M*, absence of small *m*, $FC > 2$
3. Presence of flexor *M*, presence of *CF* or *FC*, % $C > 10$, absence of object, presence of *N* or *G*
4. Presence of *M* that is neither *E* or *F*, absence of small *m*, $FC > 2$, % last 3 < 40
5. Presence of *M* that is neither *E* or *F*, and absence of object

Unimprovement:

6. Presence of extensor *M*, presence of *CF* or *FC*, % $C < 10$, presence of object
7. Presence of flexor *M*, % last 3 > 40, except when $FC > 2$
8. Presence of flexor *M*, % $C < 10$, *CF* or *FC* present, presence of object, and *N* or *G*
9. Presence of *M* that is neither *E* or *F*, presence of *CF* or *FC*, % $C > 10$, presence of object, *N* and *G* absent

II. Without regard to *M*

Improvement:

10. No small *m*, $\Sigma C > 0$, % last 3 < 40
11. Presence of *CF* or *FC*, % $C > 10$

Unimprovement:

12. No small *m*, $\Sigma C = 0$
13. % $C < 10$, presence of object, and *N* or *G*

[6] Only those patients who had at least 5 interviews are included here. Identical procedures were used with the total group with negative results.

which the most favorable results had previously been obtained in the original sample. It should also be noted that many cases do not fall under any of these rules and therefore are not judged. Results for each rule are presented in Table 5.

Table 5
Results of Empirical Pattern Analysis
(First sample $N = 50$ Improved, 59 Unimproved; Second sample $N = 70$ Improved, 34 Unimproved)

Rules*	Times cor.	Times incor.	Not judged	Times cor.	Times incor.	Not judged
Considering type of *M*						
Extensor *M*						
♯1	15	3		12	2	
(♯6)†	4	1		0	1	
			2			6
Flexor *M*						
♯2	2	0		5	0	
♯3	3	0		2	0	
(♯7)†	4	1		4	2	
(♯8)	5	2		1	1	
			11			11
M neither *E* nor *F*						
♯4	7	2		8	1	
(♯5)†	2	0		3	1	
(♯9)	5	1		1	1	
			4			10
Without regard to *M*						
♯10	4	3		16	7	
♯11	42	23		28	9	
(♯12)†	11	5		7	10	
(♯13)	11	4		9	7	
			6			11

*See Table 4 for rules. Rules are numbered in accordance with their number in Table 4.

†Numbers in parentheses refer to rules which predict unimprovement and vice versa.

The results of the cross-validation study are described in detail below and are presented in tabular form in Table 6.

Predicting Improvement: (*a*) Empirically derived rules for predicting improvement which include consideration of *M* do appear to predict improvement better than chance (see A, Table 6). (*b*) Empirically derived

Table 6
Validity of Empirical Rules

A. Pred. Imp. with *M*

	Imp.	Ui.	
With rule	30	4	34
W.O. rule	40	30	70
	70	34	104

$\chi^2 = 10.05$
$p = < .01$

B. Pred. Imp. W.O. *M*

	Imp.	Ui.	
With rule	44	16	60
W.O. rule	26	18	44
	70	34	104

$\chi^2 = 2.34$
$p = .14$

C. Pred. Unimp. with *M*

	Imp.	Ui.	
With rule	5	6	11
W.O. rule	65	28	93
	70	34	104

$\chi^2 = 2.66$
$p = .10$

D. Pred. Unimp. W.O. *M*

	Imp.	Ui.	
With rule	17	16	33
W.O. rule	53	18	71
	70	34	104

$\chi^2 = 5.47$
$p = < .02$

rules for predicting improvement which do *not* include *M* do *not* quite reach the usual criterion for better-than-chance prediction (see B, Table 6).

Predicting Unimprovement: (*a*) Empirically derived rules which include the consideration of *M* do *not* quite reach the usual criterion for better-than-chance prediction (see C, Table 6). (*b*) Empirically derived rules which do *not* include consideration of *M* *do* predict unimprovement better than chance (see D, Table 6). The above statistical presentation is modified below so that the results are made more clinically meaningful. For example, one might ask whether these data indicate when the use of *M* in clinical prediction is warranted.

Rules including M: When attempting to predict improvement using rules which include *M*, predictions could be made in 34 cases (32% of the total); these predictions proved to be correct in 30 instances (see A, Table 6). Apparently, when *M* is present this is a useful procedure. However, when attempting to predict *unimprovement* (using *M*) predictions could be made in only 11 cases (10% of total), 6 of which were correctly predicted (see D, Table 6). This does not appear to be a useful procedure. Thus, the clinical use of *M* seems indicated when predicting *im-*

provement, but not when predicting ***unimprovement.***

Rules which do not include M: When attempting to predict improvement using rules which do not include *M,* predictions could be made in 60 cases (58% of total); these predictions proved to be correct in 44 instances (see B, Table 6). This seems to be a clinically useful procedure, for the psychologist was correct about three-fourths of the time. Note, however, that the odds are biased in the psychologists' favor when predicting improvement, for the odds are about 2 to 1 that any given case in the population will be improved. It is for this reason that this rule does not quite reach the usual 5% criterion.

When attempting to predict unimprovement using rules which do not include *M,* the psychologist could predict for 33 cases (31% of the total). The psychologist was correct about 50% of the time (see C, Table 6). This hardly seems to be a useful procedure, even though better-than-chance prediction is achieved. Thus, procedures which exclude *M* from consideration seem useful clinically when predicting improvement only.

In brief, then, these data indicate that the rules for predicting improvement would be *clinically* useful, with or without the use of *M,* and that the rules for predicting unimprovement would not be clinically useful, with or without the use of *M.*[7] The reader should note again that in this population prediction of improvement is much easier than prediction of unimprovement because of the distribution of cases, i.e., there are far more improved cases--thus any prediction of improvement is more likely to be right than wrong.

Discussion

These results point to the fact that single factors on the Rorschach are not sufficient for predictive purposes. However, combination of factors, i.e., empirically determined rules for prediction, may be used with some measure of success for certain cases. The rules tested in this study were the first ones selected, and it is quite probable that they are not the best, nor the only rules which might be derived from the data. For example, shading responses were not considered nor were location responses. Other factors may well be considered. For example, we used a representative clinic sample, without regard to diagnosis or therapist. It is quite possible that if the patient population were stratified with regard to diagnosis and therapist, that greater precision might be obtained. However, the results obtained appear promising and further investigation seems appropriate.

On the other hand, the negative character of the results of our a priori procedures should not be overlooked. In this study, four procedures which might reasonably have been expected to have predictive value, did not approach success. Consider especially the "total impression" results. Here intrapsychologist agreement was higher than interpsychologist agreement. Thus, one might speculate that three rather different methods of "total impression" judgment were used rather consistently without success.

Our conclusion is that while the Rorschach test *may* be used for prognostic purposes successfully by some psychologists, there is emphatic need for further empirical check; indeed, empirical work may make prognosis from the Rorschach explicit and practical.[8]

Summary and Conclusions

1. Rorschach records of 109 unselected veterans who had been in treatment were used to predict outcome of treatment.

2. Psychologists experienced in the use of the Rorschach did not make better-than-chance predicitions of outcome of treatment under three separate methods of studying the Rorschach protocols.

3. Combinations of four variables in one instance, and five variables in another, also

[7] While differences in efficiency could be evaluated statistically, the prediction rules hardly seem well enough grounded to warrant such refinement.

[8] After this article was submitted for publication, an article by G. Filmer-Bennett entitled "Prognostic indices in the Rorschach records of hospitalized patients" appeared in *J. abnorm. soc. Psychol.,* 1952, 47, 502-506, in which the author noted that "The foregoing results failed to reveal any sharp trends in Rorschach performance which were significantly related to outcome of treatment or which compared with those reported in earlier studies."

failed to discriminate between the criterion groups.

4. The results obtained by Harris and Christianson for their group did not predict successfully for this group.

5. There was some evidence that extensor *M*'s, when they appear alone in a record, are associated with improvement in therapy.

6. Ninety-nine Rorschach "scores" were examined, and did not differentiate between the criterion groups.

7. Rules for prediction obtained on an empirical basis, and cross validated, showed definite promise of utility. The rule most useful permitted predictions in about one-third of the cases examined, and showed correct predictions at the 1 per cent level of confidence.

Received July 14, 1952.

References

1. Hammond, K. R. A tabulation method for analyzing combinations of Rorschach scores. *J. clin. Psychol.*, 1951, **7**, 276-279.
2. Harris, R. E., & Christianson, C. Predictions of response to brief psychotherapy. *J. Psychol.*, 1943, **15**, 49-74.
3. Wilkinson, B. A statistical consideration in psychological research. *Psychol. Bull.*, 1951, **48**, 156-158.
4. Zubin, J. Note on a graphic method for determining the significant difference between group frequency. *J. educ. Psychol.*, 1936, **19**, 431-444.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Rorschach Criteria for Predicting Continuation in Individual Psychotherapy[1]

Benjamin Kotkov
Boston Mental Hygiene Unit, Veterans Administration
and Arnold Meadow
University of Buffalo

This study is the second in a series of investigations designed to determine Rorschach criteria for predicting continuation in psychotherapy. The hypothesis for the study was suggested by the results of an initial study of group psychotherapy patients [3]. These results had indicated that a group of neurotic patients who continued in group psychotherapy had, on the average, higher *FC* minus *CF*, *R*, and *D%* Rorschach scores than a matched group of patients who discontinued group psychotherapy without the consent of the psychotherapist. A weighted combination of the three Rorschach scores derived by a discriminant function analysis permitted an accurate allocation of each patient to a "continuing" and "noncontinuing" group in 69 per cent of the cases. A chi-square test indicated that this percentage of successful allocation was significant beyond the 1 per cent level.

The aim of the present study was to determine at an empirical level whether the three Rorschach variables which were useful in determining continuation in group psychotherapy could predict the same behavior in individual psychotherapy.

Subjects and Procedure

The Rorschach records of twenty-six "continuing" and twenty-six "noncontinuing" patients were selected at random from the test files of the Boston Mental Hygiene Unit, Veterans Administration. A patient was classified as "continuing" if he attended the first nine psychotherapy sessions without severing his relation to the clinic. A patient was classified as "noncontinuing" if he had accepted individual psychotherapy, but ceased keeping appointments without the psychotherapist's consent before the ninth psychotherapy session.[2]

The fifty-two Rorschach records were all those of white, male veteran patients referred for individual psychotherapy. The patient group was a predominantly neurotic population with the exception of a few latent schizophrenics and behavior disorders. The age range was from nineteen to forty years with a mean age of twenty-eight years. The educational level of the group ranged from eight to fifteen years, with a mean of twelve years. There were as many married as single men. The majority were of Catholic faith. The occupations varied, with a fairly even distribution of students, clerks, skilled, semiskilled, and unskilled workmen.

The Rorshach tests had been administered by members of the Psychological staff of the Boston Mental Hygiene Unit according to the standard procedure described by Klopfer [2]. The scoring of each protocol was checked by the senior author to insure uniformity and accuracy.

The design of the experiment provides for two statistical comparisons: (*a*) between the

[1]Reviewed in the Veterans Administration and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are the results of their own study and do not necessarily reflect the opinion or policy of the Veterans Administration. The authors wish to thank Dr. Donald H. Bullock, University of Buffalo, for his help in preparing the manuscript for publication.

[2]The same number of sessions attended was used in this study as in the previous group psychotherapy study [3].

means of the continuing and noncontinuing groups on each of the three Rorschach variables, *FC* minus *CF*, *D%*, and *R.*; and (*b*) between the continuing and noncontinuing groups on a weighted Rorschach score obtained by combining the three variables according to the discriminant function equation derived from the previous group psychotherapy study.

Table 1

Means of Continuing and Noncontinuing Psychotherapy Groups for Three Rorschach Variables

Variable	Continuing group	Noncontinuing group
FC−*CF*	+.19	−1.08
D%	53.88%	45.46%
R	28.8	16.46

Results

The means of the three Rorschach variables, *FC*−*CF*, *R*, and *D%* for the continuing and noncontinuing groups of patients are presented in Table 1. The data indicate that the mean *FC* minus *CF* and *R* scores for the continuing group is higher than the corresponding means for the noncontinuing group. The tests of significance between the means yield a probability level of .01 for the *FC* minus *CF* mean difference and .02 for the *R* mean difference. The difference between the mean *D%*'s of the continuing and noncontinuing groups is not significant.

The results of the present experiment which used individual psychotherapy patients as subjects in general parallel the results obtained in the previous group psychotherapy experiment. The mean *FC* minus *CF* and *R* scores of the continuing groups of patients are significantly higher than the corresponding means for the noncontinuing groups in both experiments. In both experiments the mean *R* scores for the continuing groups are higher than the mean *R* scores for the noncontinuing groups. The test of significance for the individual psychotherapy experiment yielded a much lower probability level. The *p* for the individual psychotherapy experiment was .02; for the group psychotherapy experiment it was .01. The difference between the mean *D%* for the continuing and noncontinuing individual psychotherapy groups was in the opposite direction to that found in the group psychotherapy experiments. In the earlier experiment, however, the difference had only been significant at the .2 level.

The results of the discriminant function analysis performed in the first experiment had yielded the equation

$$z = .00241397x_1 + .00006969x_2 + .00038326x_3$$

in which x_1 is the *FC* minus *CF* score, x_2 is the *D%* score and x_3 the *R* score.[3] In the present experiment a *z* score was accordingly obtained for each patient by substituting his empirical *FC* minus *CF*, *D%* and *R* scores in this equation. The distributions of *z* scores for the continuing and noncontinuing individual psychotherapy groups are presented in Table 2. To facilitate comparison with the results of the previous experiment, the comparable distributions of the continuing and noncontinuing group psychotherapy patients are presented in the same table.

Table 2

Frequency Distribution of *z* Scores for Continuing and Noncontinuing Psychotherapy Groups

z score class intervals	Continuing individual patients	Noncontinuing individual patients	Continuing group patients	Noncontinuing group patients
1. −.00283112 – .00000000	0	3	0	0
2. .00000000 – .00283112	1	3	0	0
3. .00283112 – .00566224	3	4	7	13
4. .00566224 – .00849336	1	3	5	19
5. .00849336 – .01132448	6	8	11	17
6. .01132448 – .01415560	3	4	4	0
7. .01415560 – .01698672	3	0	9	4
8. .01698672 – .01981784	2	1	1	2
9. .01981784 – .02264896	2	0	2	1
10. .02264896 – .02548008	2	0	2	0
11. .02548008 – .02831120	0	0	0	0
12. .02831120 – .03114232	3	0	1	0
Total *N*	26	26	42	56

Table 2 indicates considerable overlap between the continuing and noncontinuing group's distribution of scores. All distributions are skewed towards the lower end of the *z*

[3]The *D%* score was multiplied by the constant 100 to eliminate the decimal point and simplify the calculations.

score range. There is, however, a greater proportion of patients with lower z scores in the noncontinuing groups than in the continuing groups in both the group and individual psychotherapy distributions.

Table 3

Test of Significance between Mean z scores of Continuing and Noncontinuing Groups for Individual Psychotherapy

	N	Mean z scores	Standard deviations	Standard errors	Critical ratio	*p*
Noncontinuing group	26	.006878	.00463	.000926	4.02	<.01
Continuing group	26	.014478	.00822	.001645		

The results of the test for the significance between the means of individual psychotherapy continuing and noncontinuing groups are presented in Table 3. Table 3 indicates that the continuing group has a higher mean z score than the noncontinuing group. Since the direction of this difference was postulated by hypothesis, a one-sided significance test is justified and the critical ratio is accordingly significant at the 1 per cent level.

Empirical inspection of the z-score frequency distributions of the group psychotherapy patients indicated that the value $z=.01132448$ represented the cutting point which produced the minimum overlap between the continuing and noncontinuing groups. In the present experiment, if we call all those cases above this point continuing and all those cases below it as noncontinuing, we find that we have predicted 36 correct placements and 16 incorrect ones. To determine whether this result could have been obtained by chance, a chi square was computed. The chi square was 7.69, a value which is significant beyond the 1 per cent level.

Discussion

An evaluation of the results of both experiments suggests the conclusion that the *FC* minus *CF* and *R* variables clearly differentiate continuing and noncontinuing patients whether in group or individual psychotherapy. The second conclusion which emerges from the data is that the *D%* is probably not a useful differentiating index in either type of psychotherapy.

When the three Rorschach variables are combined into one predictive index by a discriminant function analysis, the results for the individual psychotherapy experiment closely parallel those obtained in the group psychotherapy experiments. In both cases, the analysis yielded correct predictions in precisely 69 per cent of the cases.

In the group psychotherapy experiment, the discriminant function equation and cutting point z score were empirically derived. Accordingly, it was expected that the experimental error would reduce the accuracy of prediction in the individual psychotherapy experiment. The striking similarity of the results of the second experiment to those of the first greatly increases the degree of confidence with which the relationship, mathematically expressed by the discriminant function equation, can be accepted. However, future predictions would probably be more accurate if only the *FC-CF* and *R* variables were utilized in the equation.

The experimental results are even more impressive when certain differences between the designs of the two experiments are more closely examined. The type of subjects used in the experiments differed in many respects. In the first experiment 71 per cent of the subjects were obese nonveteran patients, and 29 per cent were neurotic veteran patients. All of the patients of the second experiment were veterans, most of whom were neurotic. Thirty-three per cent of the subjects of the first experiment were over 40 years of age. All of the patients of the second experiment were under 40 years of age. Over 50 per cent of the patients in the first experiment were women. The patients in the second experiment were all men.

Despite the differences in the types of subjects studied, the similarity of the results suggests that at least some of the personality factors which determine continuation in group psychotherapy determine continuation in individual psychotherapy.

Other personality factors not measured by the methods employed in these experiments might indicate some differential reaction

to the two types of psychotherapy. Similarly, factors other than those related to personality, such as method of psychotherapy, personality of psychotherapist, group composition, etc., may also be important determinants of whether a patient continues or discontinues psychotherapy.

The problem arises as to how to account for the results of the two experiments on a theoretical level.

It is assumed that the motivation of patients who undertake psychotherapy is ambivalent. On the one hand, they wish to continue psychotherapy in order to relieve their symptoms. On the other hand, they wish to discontinue psychotherapy because of the anxiety aroused by the psychotherapeutic process.

It is further assumed that patients with greater *FC* minus *CF* and *R* scores have a greater capacity to tolerate the anxiety aroused by the psychotherapeutic process and accordingly remain in psychotherapy. Conversely, patients with relatively low *FC* minus *CF* and *R* scores have a lesser capacity to tolerate the anxiety aroused by the psychotherapeutic process and accordingly, discontinue psychotherapy.

There are several facts which support this second assumption. Williams [5] reports a positive correlation between the number of *FC* responses and a subject's capacity to withstand experimentally induced stress. A study by Eichler [1] indicates that experimental stress produces a significant decrement in the *R* variable. It is possible that the patients with low *FC* minus *CF* and *R* Rorschach scores at the beginning of psychotherapy have initially a greater amount of anxiety than the patients with higher *FC* minus *CF* and *R* scores. They are accordingly relatively unable to tolerate the additional amount of anxiety induced by the psychotherapeutic process.

Rogers, Knauss, and Hammond [4] have recently published the results of a study similar to the one reported here. These authors attempted to determine whether the Rorschach test would aid in predicting whether an outpatient would return to the clinic for five or more interviews. On the basis of their results, the authors conclude that this behavior cannot be predicted by the Rorschach test, although they qualify their conclusion with the proviso that future, more extensively controlled studies might yield positive results. The contrasting results of the study of Rogers, *et. al.*, and those of the present study, require some theoretical interpretation.

A possible explanation can be formulated in terms of the differences in the methodology employed in the two studies. Rogers, *et. al.* used two methods for analyzing their Rorschach data: (*a*) clinical impression judgments of the entire Rorschach protocol, and (*b*) statistical comparisons of ninety-nine independent scores or signs. An objection to the use of the first method is that there was insufficient reliable information about the relationship of the independent variables (the Rorschach factors) to the dependent variable (continuation in therapy) to permit the formulation of a fruitful hypothesis. This judgment is admittedly a "Monday morning quarterback" type of criticism, but it is at least supported by the results of the Rogers study.

The sign method involves similar difficulties. The cutting points for each of the "signs" are derived on an a priori basis. The previously suggested point, that the present area of investigation is a relatively unexplored one, may also be considered as an objection to the use of a priori cutting points. In addition, the sign method is subject to the criticism that it violates the Gestalt quality of the classical Rorschach interpretive procedure.

The discriminant-function method employed in the present series of studies preserves the advantage of a method which provides for the simultaneous consideration of several Rorschach factors without sacrificing the objectivity afforded by precise quantification.

Summary

Results of a previous study of group psychotherapy patients suggested the hypothesis that continuation or noncontinuation in psychotherapy was a function of three Rorschach factors, *FC* minus *CF*, *D%*, and *R*. The aim of the present study was to test this hypothesis on individual psychotherapy patients.

Two methods were employed. The first method consisted of comparing the means on each of the three Rorschach variables of a

group of patients who discontinued individual psychotherapy with the means on the same variables of a group of patients who continued individual psychotherapy. The second method consisted of comparing the same continuing and noncontinuing groups of patients on a weighted Rorschach index. This index was obtained by the solution of a discriminant function equation empirically derived from the previous investigation of group psychotherapy patients. Twenty-six continuing and 26 noncontinuing clinic outpatients were used as subjects. The following results were obtained:

1. The average *FC* minus *CF* and *R* scores of the continuing patients were significantly higher than those of the noncontinuing patients.

2. The difference between the two groups of patients on the *D%* variable was not significant.

3. The discriminant-function-weighted Rorschach index yielded a correct prediction in 36 out of 52 cases. A chi-square test of the statistical significance of this prediction resulted in a probability value of less than 1 per cent.

The results are interpreted in terms of an hypothesis which postulates that patients with relatively high *FC* minus *CF* and *R* Rorschach scores have a greater capacity to tolerate the anxiety aroused by the psychotherapeutic process than patients with relatively low *FC* minus *CF* and *R* Rorschach scores.

Received July 22, 1952.

References

1. Eichler, R. M. Experimental stress and alleged Rorschach indices of anxiety. *J. abnorm. soc. Psychol.*, 1951, **46**, 344-355.

2. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.

3. Kotkov, B., & Meadow, A. Rorschach criteria for continuing group psychotherapy. *Int. J. Group Psychother.*, in press.

4. Rogers, L. S., Knauss, Joan, & Hammond, K. R. Predicting continuation in therapy by means of the Rorschach test. *J. consult. Psychol.*, 1951, **15**, 368-371.

5. Williams, M. An experimental study of intellectual control under stress, and associated Rorschach factors. *J. consult. Psychol.*, 1947, **11**, 21-29.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Intelligence, Verbal Fluency, and the Rorschach Test[1]

Erwin J. Lotsof
Ohio State University

It is implied in the literature that some Rorschach scoring categories have relatively distinct meanings: that is, they are related to independent personality variables. However, there is relatively little experimental work on the demonstration of this univocality. Hypotheses often engendered from a Rorschach protocol seldom can be tested or perhaps more important, they do not conform to the formal conditions of testing hypotheses. The terms in the Rorschach are rather vague, ambiguous, and almost defy being operationally defined; one need only examine words like "libidinal urges," "central anxiety," etc. "The problem facing the scientific psychologist is not, is the Rorschach valid, but, what aspects of the Rorschach are valid, and what are its limitations and uses?" [5, p. 306]. Wittenborn [9] using a factorial technique has shown that there is in fact much overlap in the scoring categories of the Rorschach using a college population. In an unpublished paper, Schroder [8] also has found much overlap in the Rorschach categories in a sample composed of individuals having Binet IQ's of 60-80 and CA's of 11.5 to 13.0.

It would appear on the basis of traditional Rorschach interpretations that scores are unrelated to the verbal fluency of the subject. From the point of view taken in this paper and implied by others [3] such a positive relation is expected to hold. In the clinical situation, one often sees the "verbal" individual produce large numbers of responses and with it a rather large number of determinants. The question then arises do these determinants actually give us the "core" of the personality or are they reflections of the patient's ability to verbalize. It was the hypothesis of this study that an explanation of the obtained overlap might be found in the possible relationship of the scoring categories to general verbal fluency. The Rorschach protocol is after all a sample of the subject's verbal repertoire. For example, it is virtually impossible to score for movement if the subject does not employ an action term in his response; likewise the scoring of shading implies the presence of certain descriptive terms in the protocol. If it were demonstrated that various Rorschach categories are not independent of each other, and if it were demonstrated at the same time that they are related to external criteria of verbal facility, then one would be tempted to question the validity of certain Rorschach interpretations. Factor analysis appeared to be the most useful technique for approaching this problem. If, as some Rorschach workers claim, the several determinants are independent, they should manifest themselves as isolated factors in a factor analysis.

Subjects and Procedure

A group of thirty college students, male and female, stratified on intelligence as measured by the Ohio State Psychological Examination, participated in this investigation. Three students were selected randomly from each decile of the OSPE. None of the subjects had taken the Rorschach prior to the present situation. The experimental situation consisted of two parts, both of which were recorded by the use of a concealed microphone. In Part 1, the subjects gave verbal descriptions of five hypothetical situations. In Part 2, each subject was administered the Rorschach test.

[1]Read in part at the Midwestern Psychological Association meetings 1951. The writer wishes to thank Dr. J. B. Rotter and Dr. R. Wherry for their aid in this investigation.

The hypothetical situations used are as follows:

1. What are the two main streets which intersect in your town? Just imagine that you are back there now on the corner. It is 12 o'clock noon on Saturday. Tell me at length everything you see and what is happening.

2. I would like you to tell me as much as you can as to what you do for recreation. Tell me all about it.

3. Suppose that you have not had a vacation for a long period of time and one day someone gives you all the money that you need to go on a trip. Tell me at length all about the trip that you would take.

4. Do you know where the oval is? Well, just imagine that you are in the center of the oval and you are telling some friend from home about *everything* you see. What would you tell them.

5. What is the easiest thing for you to talk about? Tell me as fully as possible all about it or a modification of this to fit the subject's topic.

At the termination of Part 1, the subject was told that the experimenter wanted him to do something different. At this time the Rorschach was administered employing the instructions advocated by Beck. In order to minimize the possibility of experimenter bias influencing the number of determinants, a rigid inquiry technique was employed. After the free association period, the examiner asked the following questions for each response produced:

1. Show me the - - - - - - ?

2. What gave you the idea that it was a - - - - ?

3. Can you tell me anything else that suggested that it was a - - - - - ?

After the above questions were asked, no attempt was made to probe for additional information. The Rorschachs were scored using the notations employed by Beck.

For present purposes verbal fluency was measured by counting the total number of words spoken, number of adjectives, and number of verbs.

The following criteria were used to evaluate the verbal output.

Nouns used as adjectives are counted as adjectives. Examples: *psychology* course; *toy* rifle.

Proper nouns used as adjectives are not counted. Examples: *Baker* Hall; *South* America.

Where colloquialisms or incorrect grammar are used, part of speech designated is that of the subject. Examples: It's real quiet. *Real* here is an adverb and *quiet* is the adjective.

Verbs used as nouns (gerunds) are not counted. Example: I can't get too much studying done. Studying is the gerund.

Dangling participles are counted as adjectives. *Going* into Columbus, High Street runs north and south.

Verbs used as adjectives, participles, are counted as adjectives. Example: You'd probably see people standing in groups. Standing modifies people.

Possessive forms of pronouns are not counted as adjectives, possessive adjectives.

The verb *to be* was not counted as a verb.

Auxiliaries were not counted as separate verbs. These were part of the main verb.

When standing alone they were counted.

Infinitives were not counted.

Limiting adjectives were not counted. Examples: This, those, some, many.

Ah's and similar "feeling" statements were not counted as words.

Not all of the Rorschach scoring categories were used in study. The method used in selecting the scoring categories was that these categories be relatively unaffected by the subject's verbal fluency; for example "affect" and "anxiety." The Rorschach attempts to get at basic personality variables such as "fantasy," "emotionality," "control," etc. From Rorschach literature it was felt that selecting the crucial determinants would not be related to verbal fluency. Although the superficial aspects of personality vary as a function of the situation, the core of the personality remains constant. On this basis, it would be expected that verbal fluency should be unrelated to Rorschach determinants.

Product-moment correlations were obtained among all the variables used at the present time. These variables included Rorschach scores and verbal measures. The resulting correlations were then subjected to a factor analysis using a multiple extraction process to facilitate handling of the data. The criterion of simple structure was employed in the rotation of the factors extracted.

Inspection of the correlational matrix, Table 1, shows, for this sample, disagreement with expectations derived from Rorschach theory. An interesting finding is obtained when *M* is correlated with *Sum C*. The obtained correlation of 0.54 is difficult to interpret staying within the realm of Rorschach theory; one

Table 1
Intercorrelations of Rorschach Determinants and Verbal Measures

Variable	OSPE	Total Words	Adjectives	Verbs	*M*	*R*	*W%*	*ΣC*	*A+Ad%*	*Y*	*V*
OSPE	X	.31	.33	.28	.20	.07	.15	.11	.19	.22	.33
Total Words	.31	X	.91	.97	.40	.51	.09	.51	−.04	.66	.15
Adjectives	.33	.91	X	.82	.33	.48	.05	.33	.08	.57	.15
Verbs	.28	.97	.82	X	.42	.45	.15	.56	−.01	.62	.15
M	.20	.40	.33	.42	X	.41	.14	.54	−.26	.48	.52
R	.07	.51	.48	.45	.41	X	−.29	.67	−.16	.81	.17
W%	.15	.09	.05	.15	.14	−.29	X	.13	.04	−.07	.03
ΣC	.11	.51	.33	.56	.54	.67	.13	X	−.39	.79	.30
A + Ad%	.19	−.04	.08	−.01	−.26	−.16	.04	−.39	X	−.17	−.26
Y	.22	.66	.57	.62	.48	.81	−.07	.79	−.17	X	.24
V	.33	.15	.15	.15	.52	.17	.03	.30	−.26	.24	X

might expect that the correlation should approximate zero or there should be an inverse correlation for a "normal" population. It seems necessary to conclude for this sample that there is a moderate degree of overlap between the *M* and *C* categories. Another pertinent finding is the correlation of 0.79 between *Y* and *C*. This suggests that these two determinants are not as distinct as claimed in the Rorschach literature. According to Rorschach theory, the correlation should approximate zero; it would seem rather difficult to maintain that *Y* and *C* are measuring different behaviors.

The factor analysis was based on thirty cases. With a sample of this size, the resulting factor loadings should be interpreted with some degree of caution. It is only when the loadings go above approximately .30 that some importance can be attached to them.

Table 2
Factor Matrix

Variable	Verbal Intelligence	Productivity	Elaboration	Individuality
OSPE	67	03	−02	−03
Total Words	50	34	10	06
Adjectives	47	41	02	−03
Verbs	39	32	39	05
M	30	32	12	59
R	10	96	01	00
W%	33	−31	30	09
ΣC	15	58	70	33
A + Ad%	31	29	−13	−52
Y	35	82	40	−01
V	43	15	−03	58

Four factors were obtained from a factor analysis of the correlational matrix, Table 2. The first factor has its highest loading on the OSPE with the next highest occurring on the verbal measures. The variables *Y*, *V*, *M*, and *W%* also have appreciable loadings on the first factor obtained. These are behavioral indicators which contribute to the loading of this factor but which should not do so according to Rorschach theory. The hypothesis might be ventured that the Rorschach determinants are not indicators of different behavioral manifestations but rather a priori categories which have relatively little independence. Perhaps *Y*, *V*, and *M* reflect the ability to verablize in a testing situation. If the subject cannot express himself verbally, it becomes almost impossible to score shading and vista determinants. Shading responses are supposed to indicate depression; but if the individual is depressed and refuses to verbalize, it is questionable whether or not he will produce *Y* responses. The first factor will be referred to as one of "Verbal Intelligence." Since *Y* per cent and *M* do not have high loadings on this factor, it is questionable that they are adequate indicators of intelligence. This finding is in keeping with other studies [1] which have shown that little relationship, if any, exists between *M* and intelligence.

The second factor extracted is interesting in view of the fact that some Rorschach workers feel that there is relatively high independence of Rorschach determinants; this factor will be referred to as one of "Productivity." Variable *R* has by far the maximum loading which is followed by *Y* and *Sum C*.

One conclusion appearing warranted for the sample used in this study is that the more responses the subject gives, the greater the possibility of giving some form of color or shading responses. Similar results have been obtained by Wittenborn [9] and by Schroder [8]. Again we see that Rorschach determinants have much in common; such being the case, some of the Rorschach categories might profitably be deleted or combined.

The third factor obtained is referred to as an "Elaboration" factor, that is the extent to which the individual can describe his specific perceptions. Color and shading appear to contribute a maximum to this factor. As with the previous factors we see that the color response is not a distinct category but is related to the other determinants. "Anxiety" and "affect" should not appear correlated for a "normal" population to the extent that they do in this study as well as in that by Wittenborn.

The last factor extracted is called an "Individuality" factor. Movement, vista, and color have high positive loadings while the animal percentage has a high negative loading. This factor agrees somewhat with Rorschach literature in that *M* and *A* are inversely related.

Summary

Thirty undergraduate students were administered individual Rorschachs. The subjects also gave verbal descriptions of five hypothetical situations. Intercorrelations were obtained for some of the Rorschach scoring categories and three verbal descriptive scores. The correlations were then subjected to a factor analysis. The following four factors were obtained: "Verbal Intelligence," "Productivity," "Elaboration," and "Individuality." From the loadings on some of the factors, for the present sample, it appears that Rorschach determinants are related to language measures. Specific Rorschach determinants are sometimes closely related to other determinants, in this matrix, although they previously have been considered relatively independent of the related variables. An analysis of the factors suggests that some revision is called for in the categories of Rorschach scores, and consequently the interpretations of these categories.

Received July 29, 1952.

References

1. Altus, W. D. The group Rorschach as a measure of intelligence at college level. *Amer. Psychologist,* 1949, **4,** 388. (Abstract)
2. Beck, S. J. *Rorschach's test,* Vol. I & II. New York: Grune & Stratton, 1944.
3. Eichler, R. M. Some comments on the controlling of differences in responses on the Rorschach Test. *Psychol. Bull.,* 1951, **48,** 257–259.
4. Hsu, E. H. The Rorschach responses and factor analysis. *J. gen. Psychol.,* 1947, **37,** 129-138.
5. Jessor, R. Personal communication, 1951.
6. Rorschach, H. *Psychodiagnostics.* New York: Grune & Stratton, 1942.
7. Rotter, J. B. The present status of the Rorschach in clinical and experimental procedures. *J. Pers.,* 1948, **16,** 304-311.
8. Schroder, H. Personal communication, 1951.
9. Wittenborn, J. R. A factor analysis of Rorschach scoring categories. *J. consult. Psychol.,* 1950, **14,** 261-267.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

The Effectiveness of Rorschach Elements in the Discrimination Between Neurotic and Ambulatory Schizophrenic Subjects[1,2,3]

Glenn W. Rieman
Veterans Administration Hospital, Altoona, Pa.

Since its introduction in 1921, the Rorschach test [18] has gained widespread use as a technique for the understanding and interpretation of personality. Some clinicians have developed qualitative and quantitative signs to aid in the differential diagnosis between nosological groups. Miale and Harrower-Erickson [11] proposed nine "neurotic signs" to distinguish between neurotic and normal subjects. Rorschach [1, p. 233] contended that certain signs were pathognomonic for schizophrenia. In a preliminary study, Harrower-Erickson [5, p. 114] maintained that her signs might prove to be promising in discriminating between neurotics and psychotics.

The purpose of this study was to determine whether there are statistically significant group differences between neurotic and ambulatory schizophrenic subjects with respect to certain Rorschach elements and signs. The investigation is actually a combination "element-incidence and sign" study since it involved both a preliminary search for group differences in specifically defined elements preparatory to the formulation of signs, as well as the employment of signs previously utilized. The sign approach has been used since it is an objective technique which indicates the salient features in an individual's personality structure with a minimum expenditure of time and effort. This type of investigation holds that the various components of the Rorschach test may be utilized singly or in combination, as contrasted with the intuitive or global approach alone.

Piotrowski [14, 15, 16] described ten Rorschach signs for the identification of patients with intracranial involvements. Twenty schizophrenic signs were gleaned from the literature by Klopfer and Kelley [10]. Buhler, Buhler, and Lefever [2] devised a "Diagnostic Sign List" to aid in the establishment of diagnostic-sign patterns for different clinical groups. Ross [19] applied the sign techniques of Piotrowski and of Miale and Harrower-Erickson. Davidson [4] reported a high correlation between personality patterns of children as classified by an othodox use of the Rorschach test and personal and social adjustment as ascertained by her objective Rorschach signs. Munroe [13] developed her Inspection Rorschach whereby the protocol may be scanned by means of an objective checklist in order to procure a general adjustment rating. To measure personality changes in persons undergoing psychotherapeutic treatment, Muench [12] employed 22 Rorschach factors or patterns of adjustment.

Subjects

Neurotic and ambulatory schizophrenic subjects were chosen for this study because clinicians frequently have difficulty in discriminating between them diagnostically. The neurotic

[1]This article is based upon a dissertation submitted by the writer in partial fulfillment of the requirements for the Ph.D. degree, University of Pittsburgh, 1950.

[2]Published with the permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or conclusions drawn by the author.

[3]The writer is indebted to Dr. Roy M. Hamlin, Chief Clinical Psychologist of the Western Psychiatric Institute and Clinic, University of Pittsburgh, Pittsburgh, Pennsylvania, and formerly Chief Clinical Psychologist, Veterans Administration Mental Hygiene Clinic, Pittsburgh, Pennsylvania, for a critical reading of this paper.

and the schizophrenic groups were divided into two numerically equal subgroups by the odd-even method so that there were four research groups: 50 neurotics and 50 schizophrenics for the original investigation, 50 different neurotics and 50 different schizophrenics for the replication study. Each subject chosen had a service-connected disability for either neurosis or schizophrenia. Furthermore, the diagnosis must have been confirmed currently by a psychiatrist in the installation where the patient was seen for diagnosis, treatment, or both.

It was impossible to determine whether a Rorschach test had been administered prior to the service-connected diagnosis and might have had some influence upon this diagnosis. It is recognized that most Veterans Administration installations do employ the Rorschach test almost routinely and that such results are utilized by many psychiatrists in arriving at a diagnosis. Complete control over this variable was not feasible within the scope of this study. Presumably, a patient would not be placed in either diagnostic category without one diagnosis little influenced or uninfluenced by Rorschach results. In practice, diagnoses continuing over a period of time are based primarily upon clinical behavior rather than upon test results.

Each neurotic subject had been interviewed on an outpatient basis at the Veterans Administration Mental Hygiene Clinic, Pittsburgh, Pennsylvania. Forty-four of the schizophrenic subjects were chosen from the same installation; 42 had been seen at the Veterans Administration Mental Hygiene Clinic, Newark, New Jersey; and 14 were selected from the neuropsychiatric wards of the Veterans Administration general hospitals at Aspinwall, Pennsylvania, and Butler, Pennsylvania.[4] The ambulatory schizophrenic subjects used were individuals who still maintained some contact with reality and had not deteriorated to the extent that constant hospitalization was indicated.

Only male, white, veterans of World War II were used. The schizophrenic subjects from the Pittsburgh and Newark areas were amalgamated in a random fashion; there was no significant difference between the original and the replication studies with respect to the number of Pittsburgh and non-Pittsburgh cases involved. The two nosological groups were equated on the basis of age, intelligence, and education. Ages ranged from 19 through 38. Educational levels varied from the sixth grade through one year of graduate school. Full-Scale Wechsler-Bellevue Intelligence Quotients ranged from 78 through 135. The compared groups were not at all significantly different with respect to age, intelligence, and education.

Procedure

Certain factors, ratios, patterns, and signs, designated as elements, had been compiled previously into a Rorschach Element List[5] composed of 86 items. Many of the elements tested are not signs in the strict sense of the word. A number of them are simply the various location, determinant, and content factors as used by Hertz [6] and Klopfer and Kelley [10]. Some are percentages, ratios, differences, and combinations of factors extracted from sign lists and techniques employed by Muench [12], and Klopfer and Davidson [9]. Some are qualitative characteristics which have been defined quantitatively and checked to indicate their presence. Still others are signs, in the accepted sense, which have been adopted or adapted from devices utilized by Harrower-Erickson [5], Buhler, *et al* [2], and Piotrowski [14, 15, 16].

Each Rorschach protocol was scored by the investigator from the original verbatim copy according to a prescribed procedure which may be described as a modified Klopfer system. Form accuracy, location, and populars were checked against Hertz's Tables [6]. In order to check the accuracy of the scoring, a sampling of the records was rescored independently by another experienced Rorschach

[4]Appreciation is extended to the Veterans Administration clinical psychologists who made available their psychological files and records for the accumulation and analysis of data. They are: Dr. David Cohen, Deshon Hospital, Butler, Pennsylvania; Dr. Howard W. Goodman, Aspinwall Hospital, Aspinwall, Pennsylvania; Dr. Roy M. Hamlin, Mental Hygiene Clinic, Pittsburgh, Pennsylvania; and Dr. Samuel B. Kutash, Mental Hygiene Clinic, Newark, New Jersey.

[5]The descriptive definitions of the elements used can be found in the unpublished doctoral dissertation of the writer [17] on file in the University of Pittsburgh library.

worker, using the same scoring guides, after which a percentage of agreement was calculated. Upon completion of the scoring process, the items on the Element List were tabulated for each individual record and these tabulations were amalgamated for the nosological groups under consideration. The 86 Rorschach elements were as follows:

I. Scoring Factors

1. *W*
2. *D*
3. *Dr*
4. *S*
5. *Do* (oligophrenic)
6. *M*
7. *M+*
8. *FM*
9. *m* (*Fm*, *mF*, *m*)
10. *Fk*
11. *kF*
12. *k*
13. *K*
14. *KF*
15. *FK*
16. *F+*
17. *F−*
18. *Fc*
19. *cF*
20. *c*
21. *FC′*
22. *C′F*
23. *C′*
24. *FC*
25. *CF*
26. *C*
27. *P* (popular)
28. *O* (original)
29. *H*
30. *Hd*
31. *A*
32. *Ad*
33. *At*
34. *Sex*
35. *Ar* (artificial)
36. *Na* (natural)
37. Pathognomonic content
38. Pathological content

II. Responses and Timing

39. *R* (number of responses)
40. *T/R* (average time per response)
41. Reaction time to cards without color (I, IV, V, VI, VII)
42. Reaction time to cards with color (II, III, VIII, IX, X)
43. Per cent responses to color (responses to VIII, IX, X)
44. Per cent responses to shading (responses to IV, VI, VII)

III. Form-Level Rating (FLR)

45. Total
46. Average per response
47. Average per response (noncolor)
48. Average per response (color)

IV. Patterns and Ratios

49. *F%*
50. *F+%*
51. *FM* + total *m* : *Fc* + total *c* + total *C′*
52. *A* (*A* + *Ad*) %
53. *H* + *A* : *Hd* + *Ad*
54. *W* : *M*
55. *R* + %
56. [(*Fk* + *FK* + *Fc* + *FC′*) − (*kF* + *k* + *KF* + *K* + *cF* + *c* + *C′F* + *C′*)] weighted
57. [*FC* − (*CF* + *C*)] weighted
58. *Fk* + *FK* + *Fc* + *FC′*
59. *F* : (*Fk*+ *FK* + *Fc* + *FC′*)
60. *M* + *FC* − (*CF* + *C*)
61. *M* − *FM*
62. *H* − *Hd*
63. Sum *C* (weighted)
64. *M* : Sum *C* (weighted)

V. Pathological Signs (Check presence)

65. Perseveration
66. Impotence
67. Perplexity
68. Automatic phrases
69. Color naming
70. Shading shock
71. Color shock
72. Description
73. Significant comments
74. Abstraction and symbolism
75. Confabulation
76. Contamination
77. Sex confusion or unusual handling
78. Position
79. Turning
80. Edging
81. *O* −
82. Self-reference
83. Pathological verbalization

VI. Miscellaneous

84. *"W"*
85. Rejections or failures (I through X)
86. Succession

Chi square and its probability values were utilized as the statistical tools for a test of significance of the difference between the two groups with respect to each element on the Rorschach Element List. In order to exercise some control over the possible influence of the subject's number of responses (*R*) upon certain category scores, two precautions were taken. First, the usual method of converting raw scores into percentages by dividing by *R* was employed where applicable. Secondly, an adaptation of a method advocated by Cronbach [3] was utilized to aid in partialling out differences in *R*. Each nosological group was divided at the approximate median of the *R* into two subgroups labeled "Brief" and "Longer" for separate tests of significance. If the *p* values for these two subgroups differed by less than 0.25 on each element, the "Brief" and "Longer" subgroups were ignored and only the "total" chi square and *p* value for

the element was considered. However, if the p values for these two subgroups differed by 0.25 or more, each "Brief" and "Longer" subgroup was tested separately by chi square for significant differences. In order to obtain a 2×2 table for the application of chi square, each element or element subgroup was dichotomized as near as possible to the median obtained by inspection. Elements scored for presence alone were dichotomized on the basis of presence or absence. Yates's correction for continuity [7] was used with all elements except one. Another correction was made for the few elements the cell frequencies of which were less than five. Because of small cell frequencies, several elements with similar theoretical meanings were combined; a few others could not be treated statistically for the same reason.

Since Rorschach factors are known to vary from one sample to another, further confirmation of the results of the original study was considered worth while. Those elements which revealed differences between groups at the 10 per cent level of confidence or better were singled out for additional study. The 10 per cent level was chosen as the cut-off point to allow for the inclusion and investigation of borderline factors. Those elements were retested on another sample of identical size employing the same techniques used in the original investigation.

Certain limitations of this investigation are recognized quite frankly. A major problem in dealing with nosological groups is diagnosis, especially where there is an overlapping of symptoms as in the case of neurotics and ambulatory schizophrenics. The accuracy and consistency of psychiatric diagnoses have been questioned frequently. In this study, inconsistencies of diagnosis were kept to a minimum by two procedures: (*a*) the selection of subjects was dependent upon two diagnostic judgments per individual with a time lapse between diagnoses; and (*b*) psychiatric diagnosis in both the armed forces and the Veterans Administration employs a well-standardized, tabulated, and similar plan. A second limitation lies in the problem of the homogeneity of the nosological groups under consideration since all neurotic subjects were selected from one mental hygiene clinic whereas the schizophrenics were chosen from three other installations in addition to the one just mentioned. However, both clinics are conducted under similarly oriented psychiatrists. Moreover, since deteriorated schizophrenics are not assigned permanently to the two hospitals, it would appear permissible to include schizophrenic patients from these hospitals under the general classification of ambulatory schizophrenics. A third limitation centers in the applicability of the findings. This study is concerned with groups, and any conclusions derived therefrom are applicable to similar groups alone, in the strictest interpretation. Individuals are the primary concern of clinicians. However, many persons feel that group results may serve as a base of operations for use in individual cases.

Results

With respect to the major scoring categories, the percentages of agreement on the scoring accuracy check are as follows: location, 90; determinants excluding form quality, 87; form quality of determinants, 88; content, 92; populars, 82; and originals, 33.

Table 1

Elements Showing Significant Group Differences on Both Original Investigation and Replication Study

Element	Original investigation chi square	p	Replication study chi square	p	Group exceeding on element specified
Ave. *FLR*	7.844	.002	9.000	.005	Neurotic
Ave. *FLR* (non-color)	4.000	.05	4.910	.03	Neurotic
$R + \%$	7.844	.007	5.762	.02	Neurotic
Perseveration	3.646	.06	5.018	.03	Schizophrenic
Confabulation	6.498	.01	4.158	.04	Schizophrenic
Pathological Verbalization	8.678	.006	8.384	.006	Schizophrenic

According to the established criterion, twenty elements are influenced by the number of *R*. Only 19 of the 79 elements finally tested

out in this investigation show significant differences at the 10 per cent level or better between the neurotic and the ambulatory schizophrenic groups. They are: *S* (Brief); *M%* (Longer); *FC'%* (Longer); *P%* (Total); *H%* (Longer); *T/R; RT,* Color; Average *FLR;* Average *FLR,* Noncolor; Average *FLR,* Color; Succession; *F+%*; *W*: *M* (Longer); *R* + %; *Fk* + *FK* + *Fc* + *FC'* (Longer); Perseveration; Shading Shock; Confabulation; and Pathological Verbalization.

The six elements which revealed significant group differences at the 10 per cent level of confidence or better on both the original investigation and the replication study are listed in Table 1.

The possibility that the diagnostic groups were contaminated by the use of the Rorschach in their diagnosis is a justifiable criticism and can be answered finally only by further research. A strong case can be made for the assumption that the clinical use of the Rorschach is not influenced to a major degree by the six items in the specific positive results.

Table 2

Schizophrenic Signs

Average Form-Level Rating	< 0.84
Average Form-Level Rating (noncolor)	< 0.93
R + per cent	< 82
Perseveration	Present
Confabulation	Present
Pathological Verbalization*	Present

*As defined in this study, an element which includes comments and responses that appear queer and peculiar, bizarre, incoherent and vague, illogical and absurd, critical or aggressive, self-depreciatory, irrelevant, morbid to an excessive degree, or neologistic.

While Contamination and Position did not show significant group differences on the original investigation but did so on the replication study, their comparative frequencies suggest that the presence of either type of response is practically pathognomonic for schizophrenia.

The six statistically significant elements tabulated in Table 1 were converted into the "schizophrenic signs" listed in Table 2 by means of cut-off points at the inspectional median in the case of the first three listed, while presence alone was considered a sign in the case of the last three listed. The total number of these signs for each of the individual schizophrenic and neurotic cases used in both studies was determined. These results are presented in Table 3. By means of chi square, it was determined that ambulatory schizophrenics reveal a higher number of these six diagnostic signs, statistically significant at the .001 level of confidence, than do neurotics. If the presence of any three of these six diagnostic signs were chosen arbitrarily as the cut-off point for a differential diagnosis between neurosis and schizophrenia on an individual basis, 65 per cent of the schizophrenics would have been identified properly. However, 29 per cent of the neurotics would have been included falsely within the schizophrenic classification. If five signs were selected as the cut-off point, a mere 1 per cent of the neurotics would have been categorized mistakenly but only 18 per cent of the schizophrenics would have been identified. Such results may have some value for group purposes but probably embody too much inaccuracy for individual differentiation.

Table 3

Distribution of Schizophrenic Signs Shown by Subjects in Each Nosological Group

Total signs per subject	Neurotic	Schizophrenic
6	0	3
5	1	15
4	10	18
3	18	29
2	8	11
1	23	13
0	40	11
	$N = 100$	$N = 100$

Summary and Conclusions

1. Neurotics and ambulatory schizophrenics as groups do not show significant differences in the great majority of the Rorschach elements treated. Finer tools within the elements themselves might prove more productive.

2. Five elements showed significant group differences at the 5 per cent level of confidence on both samples. They are: Average Form-Level Rating, Average Form-Level Rating (noncolor), *R*+per cent, Confabulation, and Pathological Verbalization. Per-

severation has been retained as significant even though it did not quite meet the above-mentioned standard on one of the two samples.

3. This research indicates the value of a replication study in testing Rorschach factors.

4. A worth-while by-product of this study is the demonstration that two trained individuals working independently with identical scoring guides agree to a large extent in the scoring of Rorschach responses.

5. Several observations may be made with respect to the theoretical meanings of the elements showing significant differences between the two diagnostic groups. The neurotic groups have more capacity to perceive commonly seen concepts and are better able to organize and elaborate accurately than the schizophrenic groups. The schizophrenic groups are more inclined to have difficulty in shifting their thought trends, to be unduly elaborative in an unrealistic manner, and to indulge in bizarre and illogical reasoning.

6. Kendig's "Principle of distortion of reality" [8] as crucial in the diagnosis of schizophrenia has been at least partially substantiated by the present study. These schizophrenic groups do tend to confabulate, to perseverate, to verbalize peculiarly, to perceive in an unusual fashion, and to be inaccurate in organizational and elaborative capacity more frequently than the neurotic groups.

7. Diagnostic "schizophrenic signs" obtained from the statistically significant elements may be useful for group differential purposes but are probably too inaccurate for individual diagnostic discrimination. It may be that these two groups are too ill defined as diagnostic categories to offer hope of a clear-cut differentiation between them by means of Rorschach signs.

8. Although the results of this study are predominantly negative, one observation should be made. Since the two nosological groups employed overlap in symptomatology, results which are only moderately positive or indicative of a trend are actually more significant than are apparently better results with more divergent groups.

9. It is highly improbable that the positive results obtained can be adequately explained by the possibility that the criterion groups may have been contaminated by the utilization of the Rorschach in their diagnosis. However, additional research is needed to verify this contention.

10. The comparative efficiency of this Rorschach sign technique and the global approach in the diagnostic differentiation of these two nosological groups might be worthy of study.

Received June 13, 1952.

References

1. Bochner, Ruth, & Halpern, Florence. *The clinical application of the Rorschach Test.* (2nd Ed.) New York: Grune & Stratton, 1945.
2. Buhler, Charlotte, Buhler, K., & Lefever, W. D. *Development of the basic Rorschach score with manual of directions.* Los Angeles: Rorschach Standardization Studies, No. 1, 1948.
3. Cronbach, L. J. Statistical methods applied to Rorschach scores: a review. *Psychol. Bull.*, 1949, **46**, 393-429.
4. Davidson, Helen H. *Personality and economic background: a study of highly intelligent children.* New York: King's Crown Press, 1943.
5. Harrower-Erickson, Molly R. The value and limitations of the so-called "neurotic signs." *Rorschach Res. Exch.*, 1942, **6**, 109-117.
6. Hertz, Marguerite R. *Frequency tables to be used in scoring responses to the Rorschach inkblot test* (3rd Ed.) Cleveland: Department of Psychology, Western Reserve Univer., 1946.
7. Kelley, T. L. *Fundamentals of statistics.* Cambridge: Harvard Univer. Press, 1947.
8. Kendig, Isabelle V. Rorschach indications for the diagnosis of schizophrenia. *J. proj. Tech.*, 1949, **13**, 142-149.
9. Klopfer, B., & Davidson, Helen H. *The Rorschach method of personality diagnosis: individual record blank.* New York: Rorschach Institute, 1942.
10. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* (2nd Ed.) Yonkers, N. Y.: World Book Co., 1946.
11. Miale, Florence, & Harrower-Erickson, Molly R. Personality structure in psychoneuroses. *Rorschach Res. Exch.*, 1940, **4**, 71-74.
12. Muench, G. A. An evaluation of nondirective psychotherapy by means of the Rorschach and other indices. *Appl. Psychol. Monogr.*, 1947, No. 13.
13. Munroe, Ruth L. Prediction of the adjustment and academic performance of college students. *Appl. Psychol. Monogr.*, 1945, No. 7.
14. Piotrowski, Z. A. On the Rorschach method and its application in organic disturbances of the central nervous system. *Rorschach Res. Exch.*, 1936-37, **1**, 23-40.
15. Piotrowski, Z. A. The Rorschach ink-blot

method in organic disturbances of the central nervous system. *J. nerv. ment. Dis.*, 1937, **86**, 525-537.

16. Piotrowski, Z. A. Positive and negative Rorschach organic reactions. *Rorschach Res. Exch.*, 1940, **4**, 147-151.
17. Rieman, G. W. An investigation to determine which Rorschach elements reveal group differences between neurotic and ambulatory schizophrenic subjects. Unpublished doctor's dissertation, Univer. of Pittsburgh, 1950.
18. Rorschach, H. *Psychodiagnostics.* (4th Ed.) New York: Grune & Stratton, 1949.
19. Ross W. D. The contribution of the Rorschach method to clinical diagnosis. *J. ment. Sci.*, 1941, **87**, 331-348.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Productivity on the Rorschach as a Function of Order of Presentation[1]

Louis Joseph Maradie
University of Kentucky

Increased productivity in the all-color cards (Figs. VIII, IX, X) is expected on the theory that color has a stimulating effect upon affectively labile individuals. Rorschach himself believed that there was a relationship between the number of responses influenced by color and the extent of emotional excitement although he states, "The causes of this correlation remain to be discovered" [8, p. 98].

A study reported by Dubrovner [4] in which he compared the effects of standard versus achromatic plates revealed no significant shifts in productivity upon retesting. The "color cards" (Figs. VIII-X) produced more responses than the "noncolor cards" (Figs. I-VII) regardless of the presence or absence of color. In view of this it becomes tenuous to assume that color alone is responsible for increased productivity in the last three cards.

In order to provide for controlled comparisons, retesting of the same subjects under different conditions has been deemed necessary. Rabin and Sanderson [7] retested their subjects with both standard and reversed orders of presentation and upon retesting found an increase in stereotypy ($A\%$), reduction of productivity in the reverse order condition, and a reduction of "shock" in the retest irrespective of which order was first presented. It appears that with the novelty of the Rorschach situation somewhat eliminated in all but the first acquaintance with the test, comparisons between test and retest data are incongruous.

Schachtel [9] believes that both reaction time and productivity are affected by the testee's fears, attitudes, and interests which are activated by the testing situation. It seems that memory factors would strongly influence the subject's definition of the retest situation for the challenge of the now familiar and no longer enigmatic plates is dispelled.

A common clinical observation is that the initial interview situation tends to be less productive than later ones. The subject finds himself in a novel relationship which, with the establishment of rapport, becomes decreasingly disturbing to him as a frame of reference is gradually established. In the unstructured Rorschach situation, the phenomenon is doubly significant. In empirically determining the order of the plates within the series, Rorschach predicted that failure on the first plate would rarely be encountered [8, p. 16]. He realized that early-appearing hesitancy and evasiveness could be alleviated by a relatively simple introductory stimulus card. Initial restraint and anticipatory anxiety would disappear as adaptation to the task developed.

In using the latin square, this present study is designed to circumvent the artifacts which arise by retesting the same subjects under different conditions. Its specific purpose is to investigate the possibility that irrespective of the sequential positions of the cards, later-appearing cards will produce more responses than earlier cards. At the same time, this design permits an evaluation of the productivity attributable to the cards themselves, independent of their order of appearance.

[1]This study was done in partial fulfillment of requirements for an M.S. degree in psychology. The writer wishes to express his sincere appreciation for the indispensable guidance and suggestions of Dr. Richard M. Griffith and Dr. Robert D. North, his counselors and friends. Grateful acknowledgment is also made to the student nurses of Good Samaritan Hospital, Lexington, Kentucky, who, in providing the data for this experiment, graciously gave of their time.

Procedure

From a female nurses training school, a group of 22 volunteers were obtained. Their ages ranged from 17 to 28 years with a mean chronological age of 20.2 years. Of the 22 subjects tested, 2 were omitted because of extreme overproductivity of autistic quality. In such a study dealing with a limited number of subjects, variant scores offer a definite procedural problem. Dangers of bias in omitting data must be balanced with dangers of including discrepant data which would prevent later generalization to other defined and homogeneous groups. Rorschach authorities have indicated that the response totals for normal subjects rarely exceed fifty [1, 6, 8]. The response totals for the two subjects excluded were 115 and 143 respectively — both considerably more than twice as large as the next highest score of 49 responses.

In the latin-square design as used here, systematically randomized standard Rorschach plates were presented to each subject [5, p. 302]. Ten distinct sequential orders were arranged in which each card preceded and followed every other card just once [3]. This arrangement may be made more apparent by examination of Table 1 in which each card appears once in each column and once in each row. In this square, any systematic influence resulting from the positional or sequential effect of the color cards would thus tend to be minimized.

To each of these discrete orders, a subject was randomly assigned and presented with the plates peculiar to that order. The same latin square was replicated by the random assignment of an additional subject to each of the ten orders of presentation. In this manner, twenty different subjects responded to ten dissimilar orders of presentation.

The instructions to the subjects were those suggested by Beck [2, p. 2]. After having given two or more associations for a figure, any declaration by the subject that she was through with it was accepted. Encouragement took place when less than two responses per figure were obtained but was progressively reduced until it was entirely discontinued after the fifth card.

Results

Before the data from the two separate latin squares were combined, they were tested for homogeneity of total variance by means of Bartlett's test [5, p. 195]. The resulting chi-square value of .683 was not significant and offered no evidence against the hypothesis of random sampling from a common population.

In comparing the data of the separate latin squares, homogeneity of variance was also established for the residual mean square and the sequential position mean square. The *F* values obtained were 1.04 and 1.16 respectively, which were not significant at the 5 per cent level.

After combining the data from the separate latin squares it was further ascertained that differences in productivity between the sequences or pairs of subjects responding to one particular order in the latin square were not significantly different. Examination of the agreement between the samples as shown in

Table 1

Latin-Square Design of the Order of Presentation of the Rorschach Plates

Orders	Positions									
	1	2	3	4	5	6	7	8	9	10
1	I	II	III	IV	V	VI	VII	VIII	IX	X
2	II	IV	VI	VIII	X	I	III	V	VII	IX
3	III	VI	IX	I	IV	VII	X	II	V	VIII
4	IV	VIII	I	V	IX	II	VI	X	III	VII
5	V	X	IV	IX	III	VIII	II	VII	I	VI
6	VI	I	VII	II	VIII	III	IX	IV	X	V
7	VII	III	X	VI	II	IX	V	I	VIII	IV
8	VIII	V	II	X	VII	IV	I	IX	VI	III
9	IX	VII	V	III	I	X	VIII	VI	IV	II
10	X	IX	VIII	VII	VI	V	IV	III	II	I

Table 2
Sums of Responses for Orders of Sequential Presentation (Note Table 1)

Order of Presentation	Sample 1	Sample 2	Total
1	34	43	77
2	18	29	47
3	31	40	71
4	33	22	55
5	21	23	44
6	24	26	50
7	49	23	72
8	25	27	52
9	15	23	38
10	21	37	58

Table 3
Analysis of Variance Results

	S. S.	*df*	M. S.	*F*
Independent observations:				
Order of presentation	76.32	9	8.48	1.20 (not sig.)
Residual between individuals (error)	70.60	10	7.06	. . .
Total between individuals	146.92	19		
Correlated observations:				
Cards	35.32	9	3.92	4.45**
Positions	43.52	9	4.84	5.48**
Residual from latin square	78.36	72	1.09	1.23 (not sig.)
Residual within individuals	79.40	90	.88	. . .
Total within individuals	236.60	180		
Total	383.52	199		

**Significant at 1 per cent level.

Table 2 and the insignificant F value of 1.20 in Table 3 demonstrate this fact.

Table 4 reveals that the productiveness of these 20 subjects is apparently greater in the later sequential positions. The mean number of responses in the first seven positions was 51.1 while that of the last three positions was 68.7. A t value calculated for the difference between these means was 4.00, which was significant at the 1 per cent level. Testing the significance of the mean square for total responses by positions against the residual mean square within individuals, it was found that the obtained value of F (5.48) was highly significant for 9 and 90 degrees of freedom (Table 3).

Table 4
Aggregate Sums of Responses by Positions

Positions	Responses	Positions	Responses
1	52	6	48
2	46	7	59
3	51	8	63
4	48	9	66
5	54	10	77

The other factor of interest was the effect of the cards themselves. When tested for significance against the residual mean square, the mean square for cards was also found to be highly significant with an F value of 4.45 (Table 3).

Further analysis of the effect of color upon productivity demonstrated that when the number of responses of the color cards (Fig. II, III, VIII, IX, X) was compared to those of the completely achromatic cards (Fig. I, IV, V, VI, VII), the difference between the means ($t = 1.80$) was not significant. When the number of responses to Figures I-VII was compared to those to figures VIII-X, the difference between the means was significant ($t = 2.12$) at the 5 per cent level. Examination of Table 5 shows that Card X, in contrast to Cards VIII and IX, is by far the most provocative of the series. Since Cards VIII and IX are also completely chromatic and did not elicit any noticeable increment in productivity, it appears that the configuration of Card X may be the crucial factor facilitating the subjects' responsiveness.

Table 5
Aggregate Sums of Responses by Cards in All Possible Positions

Cards	Responses	Cards	Responses
I	47	VI	57
II	55	VII	58
III	55	VIII	53
IV	53	IX	59
V	48	X	79

Summary and Conclusions

The data used in this experiment were obtained from 20 of 22 Rorschach-naive student nurses. The 2 subjects were discarded because of their extreme overproductivity. The 20 subjects were presented with ten distinct systematically randomized sequential orders of standard Rorschach plates in a latin-square design. In this manner each plate appeared in all positions and preceded and followed every other plate an equal number of times. Within the limitations imposed by the homogeneity of the sample used, examination and statistical analysis of the data revealed that:

1. Irrespective of the order of the cards, the *position* of the cards is of importance with later-appearing cards producing more responses than earlier-appearing cards.

2. Irrespective of their orders of presentation or their sequential relationships, the cards themselves differ in the number of responses elicited with Card X, because of its color or configuration, or both, evoking more responses than any other.

Thus, it is shown that the positional effect, quite apart from color, could account for increased productivity in later-appearing cards. While made suspect, it is not definitely shown that color does not also have some effect. The present writer expects to isolate this factor by the presentation of achromatic plates in the same latin-square design. If similar results are obtained in this follow-up study, it will provide conclusive evidence that the configuration of the plates and the order of their presentation are the factors responsible for increased productivity in the later-appearing cards.

Received July 30, 1953.

References

1. Beck, S. J. *Rorschach's test.* Vol. II. New York: Grune & Stratton, 1949.
2. Beck, S. J. *Rorschach's test.* Vol. I. New York: Grune & Stratton, 1950.
3. Bugelski, B. R. A note on Grant's discussion of the latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.,* 1949, **46,** 49-50.
4. Dubrovner, R. J. The effect of color on productivity and reaction time in the Rorschach test. *J. clin. Psychol.,* 1950, **6,** 331-336.
5. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1951.
6. Klopfer, B., & Kelley, D. M. *The Rorschach technique.* Yonkers, N. Y.: World Book Co., 1942.
7. Rabin, A., & Sanderson, M. An experimental inquiry into some Rorschach procedures. *J. clin. Psychol.,* 1947, **3,** 216-225.
8. Rorschach, H. *Psychodiagnostics.* Berne: Hans Huber, 1942.
9. Schachtel, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. *Psychiatry,* 1945, **8,** 419-448.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Rorschach Characteristics of Persons Scoring High and Low in Acceptance of Self[1]

Robert E. Bills
University of Kentucky

A technique called "An Index of Adjustment and Values" which is designed to give measures of the concept of self, acceptance of self, the concept of the ideal self, and maladjustment has been reported by Bills, Vance, and McLean [2]. Additional studies by Roberts [6] and Bills [3] have confirmed further the reliability and validity of the Index scores.

The present study is concerned with one of the four measures of the Index—acceptance of self. Studies by Seeman [8], Snyder [10], Sheerer [9], and Stock [11], have established that during successful psychotherapy there is an increase in positive attitude toward self or acceptance of self. Since acceptance of self has been considered a central personality variable, it follows that people who are high in acceptance of self should differ, in general, in their personality characteristics from people who are low in acceptance of self.

The Rorschach test [1, 4, 7] purports to give a description of an individual's personality. If persons who are high in acceptance of self differ in personality characteristics from people who are low in acceptance of self, and if the Rorschach is able to differentiate personality characteristics, then persons who are high in acceptance of self should have Rorschach test results which differentiate them as a group from persons who are low in acceptance of self.

Design

Twenty volunteers, female students at the University of Kentucky, were tested with the Index of Adjustment and Values and the Rorschach test. This pilot group was divided into two groups on the basis of acceptance-of-self scores above and below the mean score of the group on which the Index was standardized. The division placed 15 subjects who had scores below the mean of acceptance of self in one group and 5 subjects who had scores above the mean of acceptance of self in the other group. A comparison of the Rorschach results for the two groups suggested the hypotheses listed in Table 1.

A second group of 50 volunteer students was obtained by using as the criterion of selection, scores on acceptance of self as measured by the Index of Adjustment and Values. Twenty-five of the subjects had scores which were in excess of one standard deviation above the mean acceptance-of-self score of the standardization population of the Index (this group will be referred to as the "high scorers") and twenty-five of the subjects ("low scorers") had scores more than one standard deviation below the mean acceptance-of-self score of the standardization population. The two groups were balanced for sex and Kentucky General Classification Test scores. Each group consisted of 8 male and 17 female subjects, and a test of the significance of the difference between the means of the scores of the two groups on the Kentucky General Classification Test gave a t of .750 which showed that at the .60 level the groups were not significantly different in general ability. The groups were not significantly different in respect to the number of responses on the Rorschach, as may be seen in Table 1.

Each of the 50 experimental subjects was tested with the Rorschach, using Beck's instructions [1], by an examiner who had no knowledge of the hypotheses under investiga-

[1]This study was made possible by a grant from the University of Kentucky Research Fund Committee. The author is deeply appreciative for this aid. The author would also like to express his thanks to Glen E. Roberts who worked as an assistant on the project.

Table 1

The Hypotheses of the Study and the Statistical Results

Variable	Hypothesis	Chi square	*p*
M:Sum *C*	Above Sum *C* > *M* Below *M* ≥ Sum *C*	14.346	.001
F+%	Above *F*+% lower than below	35.280	.001
FC	Above less *FC*	9.680	.01
CF	Above more *CF*	27.000	.001
CF:*FC*	Above *CF* ≥ *FC* Below *FC* > *CF*	38.782	.001
Sequence	Above fewer cards in sequence		n.s.
A%	No difference in two groups		n.s.
F%	No difference		n.s.
R	No difference		n.s.
*FC'**	No difference	2.080	.20
M	Above less *M*	2.000	.20
Maps, etc. (evasion)	Above more evasive	2.225	.20
A+*H*: *Ad*+*Hd*	Above <2½ whole to 1 part. Below >2½ whole to 1 part.	3.125	.08
P	Above more *P*	4.023	.05
%*R* VIII–X	% smaller in above group	8.013	.01
RT Chromatic	No difference		n.s.
RT Achromatic	No difference		n.s.
RT Dif. Chr. and Ach.	Above greater difference	2.922	.10
W:*M*	Above *W* > 2*M* Below *W* ≤ 2*M*	4.667	.05
T/*R*	Above react quicker	3.945	.05
No. Turns	Above fewer turns	2.885	.10
No. Cards Turned	Above turn fewer cards	2.000	.20
M:*FM*	*FM* ≥ *M* in above *FM* < *M* in below	8.323	.01
Defense†	Above more defensive	2.885	.10
H%	No difference		n.s.
c:*k*‡	No difference		n.s.
W%	Above larger *W*%		n.s.
Obj.	No difference		n.s.
N	No difference		n.s.
Fire,etc. (aggression)	No difference		n.s.
(*H*)	Below more (*H*)		n.s.
Fm‖	Above more *Fm*		n.s.

*Includes *FC'*, *C'*, and *C'F*.
†Includes *Fc*, *c*, *cF*, *k*, *kF*, *Fk*, *K*, *KF*, and *FK*.
‡Ratio of *Fc* + *c* + *cF* to *k* + *kF* + *Fk* + *K* + *KF* + *FK*.
‖Includes *Fm*, *m*, and *mF*.

tion. The Rorschachs were then scored by a second person, not the Rorschach examiner, who had no knowledge of the subjects' classification on the Index of Adjustment and Values. The scoring system used was a mixture of Beck [1], Klopfer and Kelley [4], and Krugman [5]. Beck's system was used to score populars and location, and to determine form level. In the determination of *F*+% Beck's tables were used to determine form level, and *F*+% was calculated from the formula: Good *F* divided by Rorschach *F*. Rorschach *F* is the total number of determinants minus the color and human movement responses. All other scoring was from Klopfer and Kelley except as noted at the bottom of Table 1.

The records of two of the "high scorers" and two of the "low scorers" were selected at random and scored by two judges working independently of each other.[2] Neither of the judges was acquainted with the basis of selection of the subjects nor the hypotheses being tested. Percentages of agreement between the Rorschach scorer and the two judges were calculated for each of the four cases for the location, determinant, content, and "popular" categories. The percentages ranged from 76 to 100, and 23 of the 32 percentages were 90 per cent or greater.

Results

Table 1 presents the Rorschach variables investigated, the predicted direction of the differences of the high scorers and low scorers, and the chi squares and probabilities resulting from the calculations from double dichotomous tabulations. In these tabulations one dichotomy was formed by the independent variable, acceptance of self, and the other dichotomy by the division indicated in the prediction, or where no definite prediction was made concerning the dividing point, at the median score of all 50 cases.

Discussion

The data show striking differences in the two experimental groups. Only two of the low scorers had *F*+%'s less than the median of all 50 cases, and only two of the high scorers had *F*+%'s greater than the median. It may be concluded, therefore, that the high

[2]The author is appreciative of the assistance given by O. S. McLean and R. C. Nichols, who served as the judges.

scorers and low scorers are distinctly different in respect to $F+\%$.

The number of human movement responses was not significantly different in the two groups, although the difference approached statistical significance. A comparison of the number of human movement responses with the weighted color sum, however, shows another striking difference—the experience balances of the two groups are reversed. Twenty-two of the high scorers showed the sum of the weighted color responses greater than the number of *M* and one case had *M* equal to the sum of the weighted color responses, whereas, 16 of the 25 low scorers had the number of *M* greater than the sum of the weighted color responses. This finding leads to the conclusion that the high scorers are extratensive in their experience balance while the low scorers are introversive.

The high scorers and low scorers differed significantly in the type and number of color responses. The low scorers used fewer color responses but more "well-controlled" color represented by *FC*, whereas the high scorers used a greater number of color responses and more "poorly controlled" color represented by a predominance of *CF* and *C* responses. No "pure *C*" appeared in the records of the low scorers, but 14 of the 25 high scorers showed at least one pure *C* response. Furthermore, the high scorers showed an excess of *CF* over *FC*, and the low scorers an excess of *FC* over *CF*. Of the 25 high scorers, 24 had more *CF* than *FC* responses, but 23 of the low scorers had more *FC* than *CF*. It is concluded, therefore, that if the type and number of color responses is representative of affective life, then the high scorers have a stronger and less well-controlled affect than the low scorers.

Other variables including number of populars, the percentage of responses on the last three cards, the $W:M$ ratio, the time per response, and the $M:FM$ ratio show that at the .05 level of significance, or better, that the two groups are clearly different.

It is apparent, therefore, that in respect to certain important personality characteristics the high and low scorers, as measured by the Rorschach, constitute two distinctly different personality groups. Thus, it is concluded that the Index of Adjustment and Values is able to separate groups with different personality characteristics.

Summary

On the basis of a pilot study, 34 specific hypotheses were formulated relative to Rorschach variables present in the personalities of persons high and low in acceptance of self as revealed by the Index of Adjustment and Values. Fifty new subjects with extreme acceptance of self scores were examined with the Rorschach test. Of the 34 hypotheses 23 were confirmed, 4 were denied, and 7 remain in doubt.

It was concluded that distinct Rorschach personality characteristics distinguish subjects who are high in acceptance of self from those who are low in acceptance of self.

Received July 19, 1952.

References

1. Beck, S. J. *Rorschach's test*: I. New York: Grune & Stratton, 1949.
2. Bills, R. E., Vance, E. L., & McLean, O. S. An index of adjustment and values. *J. consult. Psychol.*, 1951, **15**, 257-261.
3. Bills, R. E. Progress of research with the index of adjustment and values. Mimeographed paper, Univer. of Kentucky, Lexington, 1952.
4. Klopfer, B., & Kelley, D. M. *The Rorschach technique*. Yonkers: World Book Co., 1946.
5. Krugman, M. Class lectures, Teachers Coll., Columbia Univer., Spring, 1948.
6. Roberts, G. E. A study of the validity of the index of adjustment and values. *J. consult. Psychol.*, 1952, **16**, 302-304.
7. Rorschach, H. *Psychodiagnostics*. Berne, Switzerland: Hans Huber, 1942.
8. Seeman, J. A study of the process of nondirective therapy. *J. consult. Psychol.*, 1949, **13**, 157–168.
9. Sheerer, Elizabeth T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.*, 1949, **13**, 169–175.
10. Snyder, W. U. An investigation of the nature of nondirective psychotherapy. *J. gen. Psychol.*, 1945, **33**, 193-223.
11. Stock, Dorothy. An investigation into the interrelations between the self concept and feelings directed toward other persons and groups. *J. consult. Psychol.*, 1949, **13**, 176–180.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Adjustment and the Discrepancy Between Self Concept and Inferred Self

A. D. Calvin and Wayne H. Holtzman
The University of Texas

One of the most widely accepted beliefs among psychologists today is that discrepancy between the self concept and objective reality is a common feature of maladjustment. Such discrepancies are said to represent a lack of insight, and a major goal of psychotherapy is to help the individual reorganize his self concept so that it will correspond more adequately with the self as viewed by others. Granted that in general the above statement may be true, it does not follow, without further experimental verification, that the relationship between such discrepancies and severity of maladjustment is as simple and direct as is often assumed. The present study was undertaken to examine the nature of this relationship in several groups of college men.

Description of the Study

Population. The population consisted of 79 male students, all the members of four fraternities, A, B, C, and D, who were residing in their respective fraternity houses at the time. Each subject was well acquainted with all the others in his group, having lived continuously in the chapter house for at least five months. The number of men in fraternities A, B, C, and D was 23, 17, 20, and 19, respectively. Groups of this size are small enough to allow extensive individual interaction and yet sufficiently large to permit adequate statistical analysis.

Collection of the data. All of the subjects were first given the Minnesota Multiphasic Personality Inventory (MMPI) in group sessions. The standard instructions for administering the test were followed [3]. At the time of testing an appointment was made to see each subject individually in order to gather data concerning his self-evaluations and his judgments of the other members in his group.

At the beginning of the individual session, each subject was interviewed briefly concerning his personal history in order to establish rapport and gain certain background information. Following the interview the subject was presented with a set of small cards, each containing the name of one of his associates and a code number for recording purposes. He was then asked to rank all the members of his group, including himself, on seven personality traits. Having the individual rank himself as well as the rest of his group made possible the study of discrepancies between the self- and group-judgments of each individual. A description of each personality trait was typed on a plain, white card to facilitate the ranking. The traits used, in the order of presentation to the subject were:

1. *Leadership.* Place the individual whom you think would be most desirable as a leader to represent the group in campus affairs at the *top*; rank the others in order beneath.
2. *Tolerance.* The general disposition of an individual to be fair and patient toward people whose attitudes are opposed to his own.
3. *Adjustment.* Well adjusted, handles personal problems well, rarely shows signs of excessive worry or anxiety, seems basically happy.
4. *Drive.* Drives himself onward, always tries hard, very ambitious.
5. *Tactfulness.* Tactful, aware of the feelings of others, keen, shows good discretion, subtle, not blunt, polished and well-mannered.
6. *Gregariousness.* Very gregarious, the "life-of-the-party" type, mixes well and seems to enjoy being around people.
7. *Social Understanding.* The needs, motivations, interest, and activities of the individuals in a group are numerous. Molded together, these diverse needs, etc., give the group a common (held-by-the-majority) set of goals, objectives, and pur-

poses. At the *top*, place the individual whom you think has the best understanding of—and works hardest to achieve—these common goals, objectives, and purposes.

These particular traits were chosen because they had been used successfully in previous studies [5, 6] employing similar ranking methods.

The interview and ranking session was conducted in complete privacy and each subject was told that his judgments would be guarded in strict confidence from the rest of the group. The subjects were informed that opportunities to discuss their own MMPI scores would be available at the conclusion of the study, an arrangement which served to keep their motivational level high throughout the study.

Method of analysis. Pooled rank scores for each individual on every trait were computed separately for each group by summing the individual ranks assigned him by his associates and placing subjects in rank order according to their pooled scores. An individual's final rank on a given trait thus represented the consensus of opinion of his associates, providing an appropriate measure of one aspect of the *inferred self*, the individual "as an informed other person sees him" [4, p. 381]. Similarly, the rank which the subject assigned to himself on that same trait provided a corresponding measure of one aspect of the *self concept*, "those parts of the phenomenal field which the individual had differentiated as definite and fairly stable characteristics of himself" [8, p. 112].

The trait description for the self-enhancing or more desirable extreme of each trait continuum had been deliberately chosen for use in this study, so that self rank which was higher than the group rank for a particular individual would be indicative of a self-enhancing discrepancy. On the other hand, a self rank which was lower than the group rank would indicate a self-depreciative discrepancy.

At first glance it would appear that a simple discrepancy score could be obtained merely by subtracting the self rank from the group rank. There are two reasons why such a method is unsatisfactory. First, rank scores have a rectangular distribution while the distribution of the underlying trait is more likely to be roughly normal. A discrepancy of, say, three ranks between self and group scores at either extreme of the rank order will represent quite a different psychological distance on the underlying trait continuum than the same numerical discrepancy in the middle range. Second, the direction and size of such a discrepancy score is partly a function of the group rank. For example, an individual who is judged by his associates to be the best adjusted person in his group can only receive a self-depreciative discrepancy score while the most maladjusted individual can only get a self-enhancing score. To the extent that the group rank and discrepancy score are correlated, any conclusion based upon an analysis of such simple discrepancy scores would be an artifact.

The first of these difficulties was largely overcome by assuming that the underlying trait distribution was normal.[1] Tables were constructed for each group for converting ranks into scale values on the baseline of the normal distribution curve so that each of the resultant distributions had a mean of 5.0 and a standard deviation of 2.0 [2, p. 248]. A discrepancy score, $(G-S)$, was computed for every individual on each trait by subtracting his converted self-score, S, from his converted group score, G.

To eliminate the spurious relationship between G and $(G-S)$, the effect of G was partialled out in the following manner: A scattergram was constructed for each of the seven traits by plotting G against $(G-S)$, using all 79 individuals. Since in every case visual inspection of the scattergram showed clearly that the regression line for predicting $(G-S)$ from G was linear with homoscedastic arrays, product-moment correlation coefficients and regression equations for predicting $(G-S)$ from G were computed. As might be anticipated from the earlier discussion, the correlations (presented in Table 1) were sufficiently high in every case to require ad-

[1]The validity of this assumption in the present instance is borne out by the fact that invariably the subject would claim that he had considerable difficulty ranking individuals in the middle ranges, but that as he moved toward either extreme the task became much easier, indicating that the underlying psychological continuum was bell shaped, probably normal.

justment of the self-group discrepancy scores. Regression equations were plotted on each scattergram and corrected discrepancy scores were determined by the following equation:

$$(G-S)_c = (G-S) - (G-S)^1$$

where

$(G-S)_c$ = corrected discrepancy score
$(G-S)$ = original discrepancy score
$(G-S)^1$ = discrepancy score predicted from regression of $(G-S)$ on G.

Table 1

Correlation Between G and $(G-S)$ for Each of the Seven Traits

Trait	r
Leadership	.71
Tolerance	.68
Adjustment	.79
Drive	.59
Tactfulness	.73
Gregariousness	.63
Social Understanding	.68

The corrected discrepancy score, $(G-S)_c$, is that part of the original discrepancy between self- and group-judgment for a particular trait and individual which remains after the spurious relationship between G and $(G-S)$ has been eliminated. A positive $(G-S)_c$ can be thought of as indicating a self-enhancing tendency greater than the average individual with a comparable G score, the larger $(G-S)_c$ the stronger the relative amount of self-enhancement. Similarly, a negative $(G-S)_c$ can be interpreted as indicative of less than average self-enhancement or even self-depreciation, as the case may be.

Four measures of the discrepancy between self concept and inferred self were derived for comparison with adjustment as measured by the MMPI. The first consisted of $(G-S)_c$ for the trait *Adjustment*. The second was a more general measure of self-enhancing tendency obtained by summing algebraically $(G-S)_c$ across the seven traits for each individual. This will be referred to as *Sum* $(G-S)_c$. The third and fourth scores, $|(G-S)_c|$ for *Adjustment* and *Sum* $|(G-S)_c|$, were similar to the first two with the exception that the direction of the discrepancy was ignored in each case. In a sense they are measures of *insight* defined simply as the degree of agreement between the inferred self and the self concept.

Results

Before attempting to study the relationships between variables derived from the group- and self-judgments and scores on the MMPI, the reliabilities of these new variables should be examined. The reliability of G, the pooled group score assigned each individual, has been reported by Greenwood [1] in an unpublished study using these same data. Reliability coefficients were computed separately for each of the four groups of subjects by dividing each group into two halves and correlating the pooled scores. As estimated by the Spearman-Brown formula the reliabilities of the final pooled scores varied from .85 for *Tolerance* to .94 for *Leadership*, highly satisfactory for the present purposes.

Although the present data permit no evaluation of the reliability of S, the self-judgment, evidence is available from an earlier study on a similar group of subjects that the reliability of self-judgments is also quite high. From data provided by Ogdon [6] in a longitudinal study of 21 men in a fraternity at the University of Texas, the present authors obtained high test-retest reliability for self-judgments of leadership. The correlation between scaled ranks over a three-week period was .97. After six weeks during which the fraternity elected new officers, calling for sharp changes in the self concepts of several individuals, the test-retest reliability was .90. This suggests that self-judgments obtained by the present methods have a high degree of reliability over a time span of at least several weeks.

Since both G and S are presumably highly reliable, it follows immediately that $(G-S)_c$ and $|(G-S)_c|$ also have sufficient reliability to warrant serious investigation of their relationships to independent measures of personality.

To have clear meaning as a derived variable, *Sum* $(G-S)_c$ should reflect a consistent tendency for an individual to enhance or depreciate himself, relatively speaking, on a variety of

personality characteristics. The degree of such self-consistency was determined by correlating the sums of $(G-S)_c$ for traits 1, 2, and 7 with the sums of $(G-S)_c$ for traits 3, 5, and 6. The choice of these particular traits was based upon a cluster analysis of the trait intercorrelations [1] and represents an attempt to make the two sets of scores as comparable as possible. The situation is somewhat analogous to that frequently encountered in test theory where over-all test reliability is estimated from the correlation between two similar halves, although in the present case the halves are only roughly parallel at best. The correlation between the two sets of scores was only .35, a low value but one which is significantly different from zero. A crude estimate of the reliability of Sum $(G-S)_c$ obtained by the Spearman-Brown formula is .56. The low degree of consistency in the discrepancies between self- and group-judgments from one trait to another may be due in large part to considerable variability in the extent to which possession of a given personality characteristic provides self-esteem.

In a similar manner, the split-half correlation for *Sum* $|(G-S)_c|$ was determined and found to be only .19, too low to justify further analysis of this variable. Moreover, the distribution of *Sum* $|(G-S)_c|$, is somewhat U-shaped due to the summing of $(G-S)_c$ without regard for algebraic sign, making comparison with other variables by correlational methods difficult to interpret.

Table 2

Correlations of MMPI Scales with Adjustment and Two Measures of Self-Enhancing Tendency

MMPI Scale	*G* for *Adjustment*	$(G-S)_c$ for *Adjustment*	Sum $(G-S)_c$
Hs	.11	—.17	—.02
D	.29*	—.42**	—.17
Hy	.11	—.03	—.10
Pd	.18	—.16	—.13
Mf	.33*	—.14	.02
Pa	.09	—.16	.03
Pt	.34*	—.44**	—.14
Sc	.30*	—.36*	—.13
Ma	.14	—.03	.11

*r significantly different from zero at .01 level
**r significantly different from zero at .001 level

After first adjusting raw scores on the MMPI for *K* and then converting to *T* scores using tables given in the MMPI manual, product-moment correlation coefficients were computed between each of the standard MMPI scales and each of three scores from the rank data: (*a*) $(G-S)_c$ for *Adjustment*, (*b*) *Sum* $(G-S)_c$, and (*c*) *G* for *Adjustment*. The group score for *Adjustment* was included because of interest in the relationship between adjustment as measured by a personality inventory and adjustment as judged by an individual's close associates. Table 2 summarizes the results of this analysis.

The pattern of relationships between scores on the MMPI and *G* is somewhat similar to that for $(G-S)_c$ in spite of the fact that *G* and $(G-S)_c$ have zero correlation as a result of the procedures used to derive $(G-S)_c$. *Sum* $(G-S)_c$, on the other hand, is not significantly correlated with any of the MMPI scales. This is probably due primarily to the unusually low reliability of *Sum* $(G-S)_c$ as a measure of self-enhancing tendency.

To the extent that *D*, *Pt*, and *Sc* are measures of maladjustment, the highly significant relationship between the MMPI and $(G-S)_c$ for *Adjustment* indicates that tendency to enhance the self is inversely related to maladjustment, the more poorly adjusted the individual, the more self-depreciative, relatively speaking, he appears. This finding is in agreement with results obtained by Sheerer working in an entirely different context. In a study of counseling interviews she reported "... a marked and fairly regular increase in the measured acceptance of and respect for self from the beginning to the end of the cases" [7, p. 175]. In each of ten cases the occurrence of self-depreciative statements decreased as the client's adjustment improved during the course of counseling.

The meaning of the relationship between *G* and the MMPI becomes clearer when the structure of the MMPI is examined more closely. In a factor analysis of 15 MMPI scales, Tyler obtained five factors, the first of which was defined chiefly by *D*, *Pt*, and *Sc*, the three scales significantly related to *G* and $(G-S)_c$ in the present study. In interpreting this factor, Tyler states, "It appears . . . to reflect a chronic state of marginal maladjust-

ment including concern with one's self. Subjects who score high on the factor would tend to have difficulties in their social contacts, possibly by reason of their dissatisfactions, worry and disgruntlement, and their complaining and pessimistic attitude" [9, p. 454]. The significant relationship between *G* and the *Mf* Scale suggests that manifestation of a feminine interest pattern in a fraternity member is sometimes viewed by his associates as a sign of maladjustment.

Using $|(G-S)_c|$ as a measure of insight with respect to adjustment two groups of subjects were formed, a *High Insight Group* consisting of 24 individuals with $|(G-S)_c|$ less than or equal to 0.5, and a *Low Insight Group* comprised of 14 men with $|(G-S)_c|$ greater than or equal to 2.0. Mean T scores for these two groups on each of the MMPI scales are presented in Table 3, together with *t* values reflecting the significance of the group differences.

Table 3

Comparison of High Insight Group and Low Insight Group on the MMPI

MMPI Scale	Mean *T* Score Low Gp.	High Gp.	*t*	*p*
Hs	56.1	49.2	2.32	.05
D	55.6	54.6	0.22	—
Hy	58.0	55.7	1.13	—
Pd	60.2	52.0	2.72	.01
Mf	66.5	60.7	2.30	.05
Pa	55.6	52.8	1.15	—
Pt	60.2	57.1	0.84	—
Sc	61.0	56.0	1.48	—
Ma	59.3	56.6	0.78	—

The mean of the *Low Insight Group* is consistently higher than that of the *High Insight Group*, sufficiently so for scores in three of the MMPI scales to prove significantly different. This result indicates that individuals who manifest poor insight regarding their own level of adjustment are more likely to be maladjusted than are those who show good insight, substantiating the claim which is frequently made that lack of insight is a sign of maladjustment.

The highly significant difference between the two groups on the *Pd* Scale is of particular interest. Although the MMPI scale designations must not be taken too literally, the frequent interpretation of a high *Pd* score as indicating an inability to profit from experience, a tendency to disregard social mores, and a general lack of insight [3] is entirely consistent with the results of this study.

Summary and Conclusions

All resident members of four fraternities were administered the Minnesota Multiphasic Personality Inventory. In later individual sessions each individual ranked all the members of his fraternity, including himself, on seven personality traits, *Leadership, Tolerance, Adjustment, Drive, Tactfulness, Gregariousness,* and *Social Understanding.* A normalized pooled rank score (G) for each individual on every trait provided an appropriate measure of one aspect of the *inferred self.* The respective self-rank (S) provided a corresponding measure of one aspect of the *self concept.* Adjusted self-group discrepancy scores, $(G-S)_c$, were used to derive four measures of the discrepancy between self concept and inferred self: (*a*) $(G-S)_c$ for *Adjustment,* (*b*) *Sum* $(G-S)_c$, (*c*) $|(G-S)_c|$ for *Adjustment,* and (*d*) *Sum* $|(G-S)_c|$. The first two served as measures of self-enhancing tendency; the latter two as measures of insight.

The major findings of this study are as follows:

1. Corrected self-group discrepancy scores obtained by the present procedures are sufficiently stable to warrant serious consideration as new measures of personality.

2. Tendency to enhance the self is inversely related to maladjustment, the more poorly adjusted the individual, the more self-depreciative, relatively speaking, he appears.

3. Maladjustment as judged within a restricted normal range by an individual's associates is directly related to maladjustment as measured by a typical personality inventory, the MMPI.

4. Individuals who manifest poor insight regarding their own level of adjustment are more likely to be maladjusted than are those who show good insight.

Received July 24, 1952.

References

1. Greenwood, C. M. Leadership in normal functioning groups. Unpublished master's thesis, Univer. of Texas, 1951.
2. Guilford, J. P. *Psychometric methods.* New York: McGraw-Hill, 1936.
3. Hathaway, S. R., & McKinley, J. C. *The Minnesota Multiphasic Personality Inventory.* New York: Psychological Corp., 1943.
4. Hilgard, E. R. Human motives and the concept of self. *Amer. Psychologist,* 1949, **4,** 374-382.
5. Holtzman, W. H. Validation studies of the Rorschach test: Shyness and gregariousness in the normal superior adult. *J. clin. Psychol.,* 1950, **6,** 343-347.
6. Ogdon, D. P. The study and prediction of attitudinal changes in a normally functioning group. Unpublished master's thesis, Univer. of Texas, 1950.
7. Sheerer, Elizabeth T. An analysis of the relationship between acceptance of and respect for self and acceptance of and respect for others in ten counseling cases. *J. consult. Psychol.,* 1949, **13,** 169–175.
8. Snygg, D., & Combs, A. W. *Individual behavior.* New York: Harper, 1949.
9. Tyler, F. T. A factorial analysis of fifteen MMPI scales. *J. consult. Psychol.,* 1951, **15,** 451-456.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Recent Research with Personality Inventories

Albert Ellis
New York, N. Y.

Although it is often assumed that projective techniques of personality evaluation have virtually monopolized the field of clinical psychology in recent years, and that paper and pencil tests of personality are rarely employed by reputable clinicians, the fact remains that far more research and theoretical papers on personality questionnaires have been published during the past six years than have been published during any equivalent period in the history of psychology. Indeed, with the possible exception of the Rorschach test, it would appear that more research and clinical use is now being made of personality inventories than of any other type of assessment instrument; and if educational, vocational, counseling, and military applications are included, then it would appear that more children and adults are now being given personality questionnaires than are being given any other type of diagnostic procedure or instrument.

Several comprehensive reviews of the literature on personality inventories have appeared during the past several years, including those by Anderson and Embree [1], Ellis [3, 6, 9], Ellis and Conrad [10], Eysenck [11], Strang [14, 15], Strang and Pansegrouw [16], Traxler [17, 18], and Traxler and Jacobs [19]. The present paper will not be another such review, but will briefly report the results of a study of the effectiveness of personality inventories when they are employed for various kinds of individual and group discriminations. In the course of this study, all research reports on personality questionnaires which appeared in the literature between January, 1946 and December, 1951 were surveyed. The results of these research reports were tabulated to determine whether, according to the criterion and level of significance stipulated by the author of each report, the inventory employed in each study did or did not significantly discriminate among the individuals or groups tested.

A survey of the recent studies of personality inventories reveals the following findings:

1. In 21 studies where intercorrelations were calculated among the subscales of different inventories to determine whether they were significantly related to each other, 20 studies showed significantly high correlations, indicating that the inventories were not measuring as many independent traits as their manuals claimed that they were.

2. In 26 studies where scores on one personality inventory were correlated with scores on another inventory purporting to measure equivalent traits, 12 correlations were found to be significantly high while 14 were not.

3. In 12 studies where inventory scores were correlated with Rorschach test results, significant relationships were found in 2 instances but were not found in the other 10 studies. Significant relationships between inventory scores and other projective technique results were found in 1 instance and not found in 3 other studies.

4. In 25 studies where the ability of subjects to fake personality inventory scores was tested, it was found that they could and did fake significantly better or worse scores in 22 instances and that they did not do so in 3 instances. In several of the studies demonstrating that personality inventory scores could be faked, it was found that special detection scales, such as the F *minus* K scale of the Minnesota Multiphasic Personality Inventory, could partially spot and compensate for the faking.

5. Results of using personality inventories for making various types of group discriminations are summarized in Table 1. In general, the data in Table 1 would indicate that the inventories significantly differentiate most oft-

Table 1
Significant Discriminations Found in Research Studies with Personality Inventories, 1946 to 1951

Discriminations Tested	Number of Studies in which Significant Discriminations Were Found	Number of Studies in which no Significant Discriminations Were Found	Total Number of Studies
Inventory scores vs. diagnostic examinations:			
Neuropsychiatric conditions	58	26	84
Psychosomatic conditions	11	2	13
Delinquent, criminal, and psychopathic behavior	15	13	28
Alcoholism	6	1	7
Predicting success in treatment	24	10	34
Total	114	52	166
Inventory scores vs. personality ratings, by associates, instructors, or counselors:			
Personality ratings	8	8	16
Sociometric ratings	5	5	10
Total	13	13	26
Inventory scores vs. behavioral characteristics:			
Extent of social participation	6	3	9
Leadership success	4	1	5
Propensity to remain healthy	5	5	10
Success in courtship and marriage	3	2	5
Tendency to become a sexual deviant	2	1	3
Tendency to have a favorable self concept	5	2	7
Extrasensory perception ability	2	2	4
Success at motor tasks	7	3	10
Tendency to take pronounced attitudes on controversial questions	4	3	7
Miscellaneous behavioral characteristics	15	19	34
Total	53	41	94
Inventory scores vs. vocational and academic test and performance results:			
Tendency to have specific vocational interests	11	5	16
Membership in different vocational groups	12	7	19
Success in actual vocational performance	18	23	41
Success in academic achievement	4	21	25
Total	45	56	101
Inventory scores vs. group differences:			
Intellectual groups	12	14	26
Educational groups	2	2	4
Age groups	6	1	7
Sex Groups	13	4	17
Socioeconomic groups	3	7	10
Ethnic groups	5	0	5
Marriage and family groups	5	5	10
Disabled and ill groups	3	7	10
College groups	6	1	7
Regional and national groups	2	2	4
Religious groups	2	2	4
Veterans groups	1	1	2
Miscellaneous groups	5	1	6
Total	65	47	112
Grand Total	290	209	499

en when used with psychodiagnostic groups and least often when used with vocational and academic groups, with their discriminative efficiency when used with other kinds of criteria or groupings falling somewhere in between.

At first blush, it is difficult to say whether the results of these recent researches with personality inventories are disappointing or not. On the one hand, they would seem to indicate that the inventories give reasonably successful discriminations when used with neuropsychiatric, psychosomatic, alcoholic, age, sex, ethnic and college groups. On the other hand, they would seem to indicate that the inventories are not measuring the independent traits they are supposed to measure; that they do not agree too well with each other nor with the results of Rorschach and projective tests; that they are easily faked; and that they discriminate unsuccessfully when used with vocational, academic, socioeconomic, and disabled and ill groups. Above all, these results would indicate that, in virtually none of the areas in which they are commonly used, do personality questionnaires *consistently* show significant group discriminations. On the contrary, different experimenters keep obtaining directly contradictory results when using the same or similar inventories.

Actually, the results shown by this survey are less favorable for personality inventories than cursory examination would indicate. Detailed analysis of many of the studies surveyed shows the same methodological failings that Ellis and Conrad [10] pointed out in a previous review. That is to say, many of the studies in which significant discriminations were obtained are suspect because they involved criterion contamination or overlap, inadequate or overly diffuse samples, set-up experiments, differential motivation on the part of the subjects, and statistical and other shortcomings. Moreover, since personality inventories are normally designed for individual diagnosis or assessment, since most of the surveyed studies are of group discriminations, and since these group results are rarely consistently discriminative, it is clear that the validity of these instruments for the main specific purpose for which most of them have been designed is, as yet, far from being solidly established.

Perhaps the main failing in many of the studies surveyed is the tendency of many authors to assume that, merely because significant *score* differences have or have not been found, significant *trait* differences have or have not been proved. One researcher after another keeps concluding that "members of the unaccepted group are more likely to have personality disturbances than members of the accepted group" [2]; or "a significant relationship was found to exist between one's personality adjustment and his occupational interests" [12]; or the like. Such conclusions, of course, will not be warranted until it is conclusively shown, beyond a shadow of a doubt, that personality scores *do* have a consistently high relationship with "personality disturbance," "personality adjustment," "neuroticism," or other traits which their authors and publishers claim they have.

The results of the present study, as well as of previous surveys made by the writer [4, 5, 6, 7, 8, 9, 10], would lead to the conclusion that although modern personality inventories, when employed for group diagnosis, have *some* degree of validity, scores obtained by individual subjects are partly or largely dependent upon (*a*) the subject's specific motivation at the time of taking the test; (*b*) the conditions under which he is tested; (*c*) his intelligence; (*d*) his degree of psychological sophistication; (*e*) his general test-taking attitudes; (*f*) his general psychological defenses; (*g*) his specific (conscious and unconscious) tendencies to lie to others and to himself; and (*h*) various other motivational and attitudinal factors. To consider these factors adequately, a good many "correction" scales or devices must be included in modern inventories—such as the various correction scales which have been and still are being added to the MMPI. So far, however, such correction scales have been shown to have only a limited effectiveness; and it may be that some of the most needed of these scales—e.g., one which would indicate the extent to which subjects unconsciously defend themselves against answering inventory questions in an objective manner—are almost impossible to devise, validate, and effectively incorporate into testing procedure because of the intrinsic limitations of personality questionnaires.

When suitable correction and qualitative analytic procedures are devised for use with

inventories, they normally add enormously to the scoring and interpretation time of the tests. Thus, utilizing a test like the MMPI today involves scoring the correction as well as the clinical scales, interpreting all the scales globally in terms of the correction factors and in relation to all the other scales, and doing a profile analysis rather than simple score interpretations.

The question therefore arises: Assuming that the newer scoring and interpretation techniques give a more valid picture of the subject's "personality" than do the older methods (an assumption which has not as yet been experimentally validated), are they worth the time and effort involved? After all, personality inventories were originally devised not because no other or better methods of assessing certain human "traits" were known during the First World War (when the Woodworth Personal Data Sheet became the first of a long line of such instruments), but because they presumably offered a *quicker* and *more convenient* method of "personality" assessment than did conventional psychiatric interviews and other diagnostic techniques. If, now, as is evident from the material in Hathaway and Meehl's monumental MMPI *Atlas* [13], the average clinical psychologist must take considerable time and trouble to master the technique of scoring and interpreting inventory results, and then must spend an hour or two scoring, graphing, and interpreting the results of each individual test given, the vaunted "advantages" of using inventories become most dubious.

Otherwise stated: since personality inventories depend, in the last analysis, on printed questions, and since virtually no one claims for them the advantage of getting at unconscious and semiconscious material which effective clinical use of interview and projective methods are generally conceded to some extent to uncover, it is difficult to see why clinicians should spend considerable time first mastering and then using these inventories. The clinical psychologist who cannot, in the time it now takes a trained worker to administer, score, and interpret a test like the MMPI according to the best recommendations of its authors, get much more pertinent, incisive, and depth-centered "personality" material from a straightforward interview technique would hardly appear to be worth his salt.

Summary

A study was made of all researches on personality inventories published between January, 1946 and December, 1951. It was found that in most cases inventory scores discriminate significantly when used with neuropsychiatric, psychosomatic, alcoholic, age, sex, ethnic, and college groups. On the other hand, it was found that in most instances the inventories are not measuring the independent traits they are supposed to be measuring; they do not agree too well with each other nor with the results of Rorschach and projective tests; they are easily faked; and they usually do not give significant group discriminations when used with vocational, academic, socioeconomic, and disabled and ill groups. It was especially found that, in none of the areas in which they are commonly employed, do personality inventories consistently show significant group discriminations. It was pointed out that when these inventories are most effectively used, they tend to become equally as time consuming as alternative psychodiagnostic procedures, such as interviewing and the use of projective techniques, that seem to be more clinically incisive and valuable.

Received June 9, 1952.

References

1. Anderson, G. V., & Embree, R. B. Appraisal of the individual. *Rev. educ. Res.*, 1948, **18**, 157-174.
2. Brody, D. S. A genetic study of sociality patterns of college women. *Educ. psychol. Measmt*, 1951, **10**, 513–520.
3. Ellis, A. The validity of personality questionnaires. *Psychol. Bull.*, 1946, **43**, 385-440.
4. Ellis, A. A comparison of the use of direct and indirect phrasing in personality questionnaires. *Psychol. Monogr.*, 1947, **61**, No. 3 (Whole No. 284).
5. Ellis, A. Discussion of Heinlein's comment on "The validity of personality questionnaires." *Psychol. Bull.*, 1947, **44**, 83-86.
6. Ellis, A. Personality questionnaires. *Rev. educ. Res.*, 1947, **17**, 53-63.
7. Ellis, A. The value of marriage prediction tests. *Amer. sociol. Rev.*, 1948, **13**, 710-718.
8. Ellis, A. The relationship between personality inventory scores and other psychological test results. *J. soc. Psychol.*, 1948, **28**, 287-289.

9. Ellis, A. Self-appraisal methods. In L. E Abt, & D. Brower (Eds.), *Progress in clinical psychology*. New York: Grune & Stratton, 1952.
10. Ellis, A., & Conrad, H. S. The validity of personality inventories in military practice. *Psychol. Bull.*, 1948, **45**, 385-426.
11. Eysenck, H. J. Personality tests: 1944-1949. In G. W. T. H. Fleming, *Recent progress in psychiatry*. London: Churchill, 1950. Vol. II., pp. 197-253.
12. Feather, D. B. The relation of personality maladjustments of 503 University of Michigan students to their occupational interests. *J. soc. Psychol.*, 1950, **32**, 71-78.
13. Hathaway, S. R., & Meehl, P. E. *An atlas for the clinical use of the Minnesota Multiphasic Personality Inventory*. Minneapolis: Univer. of Minnesota Press, 1951.
14. Strang, Ruth. Methods, technics, and instruments of mental hygiene diagnosis and therapy. *Rev. educ. Res.*, 1946, **16**, 436-443.
15. Strang, Ruth. Technics and instruments for the improvement of mental health. *Rev. educ. Res.*, 1949, **19**, 386-394.
16. Strang, Ruth, & Pansegrouw, Deborah. Studies of individuals. *Rev. educ. Res.*, 1948, **18**, 382-395.
17. Traxler, A. E. Measurement in the field of personality. *Education*, 1946, **66**, 424-430.
18. Traxler, A. E. Nonprojective technics in the appraisal of personality. *Growing points in educational research*. Washington: American Educational Research Association, 1949. Pp. 324-330.
19. Traxler, A. E., & Jacobs, R. Construction and educational significance of structured inventories in personality measurement. *Rev. educ. Res.*, 1950, **20**, 38-50.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Distortion in Paranoid Schizophrenia[1,2]

Paul G. Daston, Gerald F. King
VA Hospital, Ft. Custer, Michigan, and Michigan State College

and Stewart G. Armitage
VA Hospital, Ft. Custer, Michigan

On the basis of clinical observations and experimental evidence schizophrenic thought processes have been described as being rigid, concrete, and distorted. However, experimental studies in this area have not provided consistent support for most of these generalizations. The purpose of this research was to investigate distortion in schizophrenia using a controlled quantifiable approach.

In reference to schizophrenic behavior, what might be called distortion has been described in many ways, ranging from Beck's [3] "inaccurate perception" of clearly implied forms to Cameron's [4] "displacement" of precise verbal terms by approximate concepts. In Aborn's review of the literature, he concluded that although "it may be regarded as secondary to some more fundamental notion of abnormality, distortion is in its broadest sense primarily descriptive of the introduction of contradictory elements into a situation providing no external barriers to consistent behavior" [1, p. 7].

Methodology

The methodology employed in this research stems from the work of Asch [2], Goodman and Lorge [5], and Aborn [1]. Using descriptions of hypothetical people, Asch was interested in how an over-all impression of a person was formed. From the responses of his subjects to a checklist of trait names, he found that an impression of personality is a function of an interaction of traits rather than of simple summation. A variation of this procedure was developed by Goodman and Lorge and utilized in a study of the dynamics involved in the formation of personality impressions by seventh- and ninth-grade school children. Aborn modified the technique as it is used in the present research.

Studying distortion in schizophrenia, the latter investigator constructed a number of short stories describing hypothetical people. Some of the stories utilized positive trait adjectives only, and some contained only negative traits. Further, one set of descriptive stories explicitly stated trait names, while another simply implied them. For example, the all-positive story with trait names explicitly stated was as follows:

> The first things people notice about him are that he is cheerful and friendly. In school, he always got good marks and was looked upon as being intelligent. He is very active and takes part in all sports. As a player, he is known to be fearless and is respected because he is honest.

In comparison, the all-positive story with the traits implied was:

> On his way home, he found a pocketbook in the street and turned it over to the policeman on the corner. He smiled and greeted most of the people he passed along the street and stopped to talk with the man at the newspaper stand. The newspaper man likes to listen to him because he usually knows what is going on in the world. Sometimes they talk about baseball, since they both play on the neighborhood team. They have known each other ever since he chased and caught two men who had robbed the newspaper stand.

[1]This research was presented in part at the meetings of the Midwestern Psychological Association in Cleveland, April, 1952.

[2]Published with permission of the Chief Medical Director, Department of Medicine and Surgery, Veterans Administration, who assumes no responsibility for the opinions expressed or the conclusions drawn by the authors.

The authors are especially indebted to the cooperation and assistance provided by Dr. Murray Aborn of Michigan State College.

Immediately after reading each story, the subject indicated his "picture" of the person in the story by checking appropriate words on a word-selection checklist, which consisted of 36 positive and 36 negative adjectives. The checklist was constructed to sample all areas of personality description, with the affective tone of the adjectives being based on the ratings of 62 judges. For each story, the subject's responses could be quantified in terms of *return of givens, expansions,* and *distortions.* A distortion was operationally defined as a response which was inconsistent with the given trait pattern of a story; for example, checking a negative adjective for a positive story.[3]

While Aborn's results failed to show any statistically significant differences between a mixed group of schizophrenics and normals, a trend toward greater distortion on the part of the experimental group was noted. It seemed possible that this trend might be accentuated if more opportunity were allowed for the schizophrenics to introject the stimulus material by inserting a time interval between the presentation of the stories and the checklist.

Experimental Procedure and Subjects

In testing this hypothesis, the following two modifications in methodology were employed: (*a*) the experimental group was confined to paranoid schizophrenics; and (*b*) a temporal factor was introduced. After reading each story, the subject spent five minutes on a simple motor task in which he was required to cross out *O*'s and circle *X*'s. Then he was presented with the checklist and asked to check all the words he would use to describe the person in the story. This procedure was followed for all four stories. The stories were presented in a different order to different subjects to minimize any transfer effects that might be involved.

The paranoid schizophrenic and normal groups both contained 27 male subjects. The groups were equated for age, education, and intelligence, as shown in Table 1. The subjects were selected from a rather limited age range; it actually extended from about 22 to 35 years. The measure of intelligence was a short form of the Wechsler-Bellevue Scale, Form 1 (consisting of the Comprehension, Arithmetic, and Similarities subtests), which Rabin [6] and others have shown correlates very highly with the full scale with both normals and neuropsychiatric patients. In order to reduce the possibility that some of the words in the stories or checklist might be misunderstood by the subjects and thus contaminate the results, the minimum IQ established for the selection of subjects was 100. Further, since none of the words in the stories and checklist were, according to the Thorndike-Lorge word count [7], above seventh-grade comprehension (an educational level below all the subjects), this was deemed an adequate safeguard. An attempt was also made to select the two groups so there would be an equivalent range of occupations.

Table 1

Group Characteristics of the Paranoid Schizophrenics ($N = 27$) and the Normals ($N = 27$)

Group	Age		Education		Intelligence IQ	
	Mean	*SD*	Mean	*SD*	Mean	*SD*
Normals	29.6	4.0	12.3	2.5	115.8	10.0
Schizophrenics	30.6	5.9	12.0	2.6	116.1	12.2

The subjects of the experimental group had all been given a final diagnosis of "schizophrenia, paranoid type" by a staff of hospital psychiatrists. The majority were open-ward patients, and all were cooperative and in good contact. All the normals were gainfully employed, had never been hospitalized for neuropsychiatric reasons, and were not currently undergoing treatment.

Results

A comparison of the two groups on the total number of adjectives used for each story, summarized in Table 2, reveals no differences. None of the computed *t* ratios are above 1.00. The data were then analyzed on the basis of return of givens for each story, which refers to the checking of trait adjectives stated directly or strongly implied in the stories. As can be seen in Table 3, there was a tendency for the

[3]For a more detailed account of Aborn's methodology, a microfilm of his complete manuscript can be obtained from University Microfilms, Ann Arbor, Mich. Publ. No. 2336, Pp. 84.

Table 2
Comparison of the Experimental and Control Groups on the Total Number of Adjectives Used

Group	Total Number of Adjectives Checked Stories*							
	1–A		2–A		1–B		2–B	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Normals	24.52	8.69	17.78	8.07	24.59	9.37	18.11	6.70
Schizophrenics	26.81	8.56	20.37	10.41	23.70	8.85	19.37	8.01

*Key: 1, all positive traits; 2, all negative traits; A, traits stated directly; B, traits implied.

Table 3
Comparison of the Experimental and Control Groups on Return of the Givens

Group	Return of Givens Stories							
	1–A		2–A		1–B		2–B	
	Mean	*SD*	Mean	*SD*	Mean	*S*	Mean	*SD*
Normals	5.33	0.94	4.63	1.39	5.11	1.02	4.41	1.49
Schizophrenics	5.00	1.19	4.11	1.81	4.70	1.24	3.78	1.90
t ratio	1.14		1.18		1.29		1.33	

normals to return more givens and the schizophrenics to be more variable. Neither trend, however, reaches statistical significance.[4]

In analyzing the distortion scores, the chi square yielded by Bartlett's test of homogeneity showed that homogeneity of variances for the two groups could not be assumed, which introduces certain difficulties for standard statistical treatment. An inspection of the data suggested that the extreme scores of six schizophrenic subjects (high in distortion) accounted for the marked variance of the experimental group. When the extreme distortion scores made by these subjects (a total of 11 scores) were eliminated and the chi square was computed again, homogeneity of variances could then be accepted. We now felt justified in employing analysis of variance since discarding these extreme scores depressed the mean distortion values for the experimental group, thus working against our hypothesis. As such, it would constitute a conservative or "weak" test of significance.

The results of the analysis of variance are presented in Table 4. It can be seen that the paranoid schizophrenics produced significantly more distortion than the normal controls (beyond the 1 per cent level of confidence). Comparing the groups on individual stories showed significant differences in stories 2-A (1 per cent level) and 1-B (5 per cent level) which does not suggest any particular pattern. The interaction term between the diagnostic categories and stories was not significant, indicating the experimental subjects did not react to the materials in a different manner than did the controls. It was a reaction of greater magnitude but in the same direction.

Table 4
Treatment of Distortion Scores by Analysis of Variance

Source of Variation	*df*	Sum of Squares	Mean Square	*F*	*p*
Total	204	282.71			
Diagnostic Categories	1	13.74	13.74	19.35	<.01
Stories	3	17.81	5.94	8.37	<.01
Individuals	53	144.87	2.73	3.85	<.01
Stories vs. DC	3	4.56	1.52	2.14	NS
Error	144	101.73	0.71		

In summary, the paranoid schizophrenics showed both significantly more distortion and

[4]These results coincide with the findings of Aborn who used return of givens as one index of rigidity.

more variability than did the normals. It should be pointed out, however, that this difference in amount of distortion is a group difference, with considerable overlapping between groups. As such, it is not adequate for individual prediction.

Discussion

In light of Aborn's findings, it would appear that schizophrenics do not necessarily deviate perceptually from normals in the immediate recall of controlled verbal material. However, when a temporal interval is interpolated between presentation of the material and recall, distortion occurs. This is seemingly a result of the interaction of a more or less correct perception with the intervening schizophrenic personality variables. The results suggest that it is the processes underlying schizophrenic thought or judgment that somehow go awry.

It is realized that there are certain difficulties in comparing the findings of this research with those of Aborn. Since two modifications in his methodology were used, the question arises as to which variable accounts for the distortion. The interpolated time interval has already been discussed. The other possibility lies in the nature of the subjects used. Although Aborn's group of schizophrenics of various subtypes was comparable to that of this research in terms of such characteristics as age, education, and intelligence, our experimental group was confined to paranoid schizophrenics. Since the principle criterion in formulating the diagnosis of "paranoid schizophrenia" is the presence of "projective" thinking (as seen in delusions), this factor may be the crucial one in producing distortion. The answer to this question seems obtainable in further research with this technique in which the following comparisons would be made: (*a*) paranoid schizophrenics vs. normals without the temporal interval; and (*b*) nonparanoid schizophrenics vs. normals with the temporal interval.

Summary

The purpose of this research was to investigate distortion, a characteristic often attributed to the behavior of schizophrenics. The methodology employed four short stories describing hypothetical people and a word-selection checklist. After reading each story, the subject spent five minutes on a simple motor task. He was then presented with the checklist and asked to check all the words he would use to describe the person in the story. A distortion was defined as the checking of a word which was inconsistent with the given trait pattern of a story. The subjects consisted of 27 paranoid schizophrenics and 27 normals, the two groups being equated for age, education, and intelligence. A statistical analysis of the data revealed that the experimental group showed both significantly more distortion and more variability than the normals. The results were compared and discussed in terms of the findings of Aborn, who developed the basic methodology used in this research.

Received June 23, 1952.

References

1. Aborn, M. A study of rigidity and distortion in normals and schizophrenics with controlled verbal material. Unpublished doctor's dissertation, Columbia Univer., 1950.
2. Asch, S. E. Forming impressions of personality. *J. abnorm. soc. Psychol.*, 1946, **41**, 258-290.
3. Beck, S. J. Personality structure in schizophrenia. *Nerv. ment. Dis. Monogr.*, 1938, No. 63.
4. Cameron, N. Reasoning, regression and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
5. Goodman, S. Forming impressions of persons from verbal report. Unpublished doctor's dissertation, Columbia Univer., 1950.
6. Rabin, A. I. A short form of the Wechsler-Bellevue test. *J. appl. Psychol.*, 1943, **27**, 320–324.
7. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Bur. Publ., Teachers Coll., Columbia Univer., 1944.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Some Effects of Sodium Amytal on "Deteriorated" Schizophrenics[1]

Willard A. Mainord
State College of Washington[2]

Conflict still exists as to the nature of schizophrenia, much of the controversy centering upon whether this pathology is of an organic, functional, or psychosomatic nature. Schizophrenic "deterioration," in particular, remains obscure as to nature and genesis, and it is obvious that more knowledge is necessary for best clinical results. Three types of theories are advanced to explain schizophrenic deterioration, i.e., schizophrenia is a progressive, organic disease; schizophrenia is a strictly functional disorder which serves some defensive function in the psychological economy of the individual; and, schizophrenia is a psychosomatic malady, and causal factors as well as disease processes may be both organic and functional. Cameron [2] notes that a rigid organic point of view usually implies actual brain deterioration not amenable to anything but physiological treatment, and with little hope of full restitution of intellectual efficiency.

The usual definition of schizophrenic deterioration is descriptive, but recently Levy and Kennard [7] have advanced a concept of deterioration with more objective criteria. Their study shows a relationship between clinical judgment of deterioration, abnormality of EEG records, and significantly large Wechsler-Bellevue deterioration ratios. These criteria can be duplicated by other investigators, and comprise the operational definition of schizophrenic deterioration used in this paper.

If schizophrenic deterioration can be alleviated without organic change, psychological therapies would appear to offer hope for rehabilitation of deteriorated patients. Layman [5, 6] has demonstrated that it is possible to obtain a temporary improvement in the intellectual performance of an unselected group of schizophrenics through injections of sodium amytal. There is evidence [1, 8, 9] which indicates that these results may come from changes in motivational states which come about as a result of the action of this drug. The problem investigated in this paper can be stated as follows: can a statistically significant improvement in performance be obtained from deteriorated schizophrenics, as defined by the criteria suggested by Levy and Kennard, while the patients are under the influence of sodium amytal? The working hypothesis was that such an improvement could be obtained because of motivational changes resulting from the drug.

The nature of the criteria suggested the following secondary hypotheses: if schizophrenia is a progressively disabling disease, there should be a significant negative correlation between the subjects' ages and their obtained IQ's; and there should be a significant positive correlation between the improvement observed under amytal and the size of the deterioration ratio. This latter hypothesis is based on the assumption that the Wechsler-Bellevue deterioration ratio does measure organic damage, and that the greater the amount of tissue pathology, the greater the amount of potential improvement.

[1]This paper is derived from a study done as a master's thesis under the direction of Dr. J. H. Elder to whom grateful acknowledgment is made. Acknowledgment is also made to Dr. Sol Levy, Clinical Director, Eastern State Hospital, Medical Lake, Washington and to the clinical staff for the interest and cooperation shown relating to this study. All conclusions, of course, are the author's own, and the responsibility for them entirely his.

[2]Now at the University of Washington.

Procedure

The criterion given the most weight in the selection of subjects was clinical judgment, inasmuch as behavior itself is basic to any concept of deterioration. As an additional precaution, the theorized progressive nature of the disease was allowed for by selecting only subjects who were considered to have had the illness for at least five years. The abnormality of the EEG records was determined by a competent specialist. The minimum Wechsler-Bellevue deterioration ratio acceptable for the experimental group was 15 per cent after correction for aging—as Wechsler [11] suggests 10 per cent as a minimum significant figure. All subjects who had been subjected to any kind of therapy after the original testing were excluded. Table 1 summarizes the mean age, known duration of illness, deterioration ratio, and IQ for both the experimental and test-retest groups. The test-retest group was used only to give some indication of usual test-retest effects on long-standing schizophrenics, inasmuch as the experimental control was accomplished by comparing pre-amytal with post-amytal scores. All cases were selected from the same population as that used by Levy and Kennard.

Table 1

Mean Age, Known Duration of Illness, Deterioration Ratio, and IQ for Experimental and Test-Retest Groups.

Group	*N*	Mean Age	Mean Duration	Mean Deterioration ratio*	Mean IQ*
Experimental	22	43.5	13.6	31.4%	84.5
Test-retest	18	41.0	10.4	12.6%	100.6

*The utilization of the criteria for selection so diminished the original population that it was impossible to match these two groups for IQ and for deterioration ratio.

The subjects, experimental and test-retest, were originally tested with the Verbal section of Form I of the Wechsler-Bellevue Intelligence Scale in a two-week period from December 20, 1950 to January 5, 1951. The experimental group was retested approximately nine months later while under the influence of sodium amytal.[3] The final test was administered to the experimental group approximately one week later, again under conventional conditions. The test-retest group was also retested at this time.

Results

The results obtained can be said to bear out the hypothesis that a significant, temporary improvement in measured intellectual performance would be obtained through the administration of sodium amytal.

Table 2

Mean Group Scores of Experimental and Test-Retest Groups on Wechsler-Bellevue Verbal Scale under all Test Conditions

Group	Pre-Amytal	Under Amytal	Post-Amytal
Experimental	84.5	94.1	86.4
Test-retest	100.6	—	101.6

Table 2 summarizes the results, in which it can be seen that the mean Verbal IQ of the experimental group increased to 94.1 from 84.5 under sodium amytal, and dropped back to 86.4 in the post-amytal testing. In other words, the experimental group, while under the influence of the drug, increased its mean Verbal IQ by 9.6 points (significant at better than the .001 level of confidence), and then lost 7.7. points of this gain (significant at better than the .001 level of confidence) when the same test was administered again without the drug. The difference between the means of the pre- and post-amytal conditions was not statistically significant, despite practice effects. The test-retest group improved its performance from 100.6 to 101.6, this one-point gain being comparable to the 1.9-point gain between the first and third testing of the experimental group.

The secondary hypotheses were not confirmed by the data, inasmuch as the correlations obtained in each case were insignificant. Thus, only the primary hypothesis, i.e., that a

[3]The drug was administered intravenously in a solution composed of one grain per cc. of distilled water. No set dosage was used; the drug was discontinued when the subject seemed to have reached a testable level which appeared optimal from observer judgment.

significant improvement in performance of deteriorated schizophrenics could be obtained by injection of sodium amytal, was supported by the results obtained.

Discussion

If sodium amytal does not temporarily mitigate organic pathology, the obtained data would seem to support the point of view that schizophrenic deterioration is not entirely physiological. In view of other work [8, 9], it would seem logical to ascribe the improvement noted to changes in the patients' motivation. Therapeutically, it might be expected that schizophrenia and its accompanying deterioration may be partly mitigated by psychological methods. A permanent improvement in all spheres of behavior of comparable magnitude to the temporary intellectual improvement obtained in this study would be enough to return many deteriorated schizophrenics to society. Failure of such methods to date may be due to lack of perseverance with difficult patients, inadequate techniques, and lack of funds at the insistence of economy-minded legislators and administrators.

This study suggests that the severity of the schizophrenic process may have little or nothing to do with length of illness. A familiar sight in any institution is the well-preserved paranoid schizophrenic of many years standing. This study further suggests that no matter what Wechsler's deterioration ratio does measure [3, 4, 10, 11], it does not seem applicable to schizophrenic deterioration, as no relationship between these ratios and the data could be detected.[4]

Of striking interest, but difficult to evaluate, were the results obtained through the use of amytal with patients otherwise untestable. Five such patients (none of them otherwise used in this study) were tested, and four of them yielded scorable protocols with Verbal IQ's ranging from 60 to 95, apparently indicating that much potential is still available to many of the most deteriorated patients.

[4]Subtest analysis was also sterile both in terms of isolating differential improvement and in terms of subtest scatter. Amytal improvements were evenly distributed over all of the subtests, and the scatter remained extensive—often increasing.

Summary

Twenty-two deteriorated schizophrenics were tested to obtain psychological measurement of intellectual ability both with and without the use of sodium amytal. Deteriorated schizophrenics were defined as those: who had a known illness of at least five years duration; who were judged clinically to be deteriorated; who showed a significantly large Wechsler-Bellevue deterioration ratio; and who showed an abnormal EEG record. A test-retest group, comparable in age, length of illness, and clinical judgment was used to obtain some data on test-retest effects.

The experimental group was administered the Verbal Scale of the Wechsler-Bellevue three times—the first time conventionally, the second time under sodium amytal, and the third time under usual conditions. The test-retest group was conventionally tested twice. The primary hypothesis tested was that a statistically significant improvement in performance would be obtained with subjects under sodium amytal over non-amytal tests. This hypothesis was confirmed by the data, as these differences were statistically reliable, while the differences between the first and third testings were statistically insignificant.

A discussion of implications for psychotherapy followed indicating that the obtained results support the argument that psychotherapy might be successfully utilized with deteriorated schizophrenics, and that such patients still have a potential capacity beyond their observed behavior. The applicability of the Wechsler-Bellevue deterioration ratio to this illness was considered, the conclusion being that such use was of extremely doubtful validity. The over-all conclusion from this study was that there is a functional component in schizophrenic deterioration which may be amenable to psychological methods.

Received July 28, 1952.

References

1. Arkola, Audrey L. An experimental study of the effects of sodium amytal upon performance of the Hunt-Minnesota Test for organic brain damage. *J. clin. Psychol.*, 1947, **3**, 392-396.
2. Cameron, N. The functional psychoses. In J. McV. Hunt (Ed.), *Personality and the behav-*

ior disorders. New York: Ronald, 1944. Pp. 861-921.

3. Garfield, S. L., & Fey, W. F. A comparison of the Wechsler-Bellevue and the Shipley Hartford Scales as measures of mental impairment. *J. consult. Psychol.*, 1948, **12**, 259-264.
4. Hunt, W. L. The relative rates of decline of Wechsler-Bellevue "hold" and "don't hold" tests. *J. consult. Psychol.*, 1949, **13**, 440-443.
5. Layman, J. W. A quantitative study of certain changes occurring under sodium amytal. *Psychol. Bull.*, 1936, **33**, 745.
6. Layman, J. W. A quantitative study of certain changes in schizophrenic patients under the influence of sodium amytal. *J. gen. Psychol.*, 1940, **22**, 67-86.
7. Levy, S., & Kennard, Margaret. Unpublished study.
8. Lindemann, E. The psychopathological effect of sodium amytal. *Proc. Soc. exp. Biol. Med.*, 1931, **28**, 864-866.
9. Lindemann, E. Psychological changes in normal and abnormal individuals under the influence of sodium amytal. *Amer. J. Psychiat.*, 1932, **11**, 1083-1091.
10. Magaret, Ann, & Simpson, Mary M. A comparison of two measures of deterioration in psychotic patients. *J. consult. Psychol.*, 1948, **12**, 265-269.
11. Wechsler, D. *The measurement of adult intelligence.* (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Influence of Socioeconomic Status on Wechsler Intelligence Scale for Children: An Exploratory Study

Betsy Worth Estes
University of Kentucky

The present study is a preliminary exploration toward determining whether socioeconomic level affects responses to the items of the Wechsler Intelligence Scale for Children so that test differences between upper and lower socioeconomic levels are manifest in:

1. The distribution of IQ's (means and variances).

2. The prediction value of the WISC, i.e., its correlation with an achievement test.

3. The pattern of test responses as indicated by subtests and intercorrelations between the subtests.

Subjects and Procedure

Two groups of children were studied: one of upper economic status (N=40), one of low economic status (N=40). The subjects in the upper economic group were from the university school, which is essentially a public school except for the nominal tuition charged. This school is one of several in which can be found a fairly large number of children from very comfortable homes—upper economic status. The fathers of these children are university professors, lawyers, physicians, bankers, editors, merchants, clergymen, etc. The subjects in the low economic group were from five public schools. The fathers of these children are day laborers, night watchmen, junk collectors, etc.

Of the total of 80 children, one-half were boys and one-half were girls. One-half were seven years of age, plus or minus two months, and in the second grade; the other half were ten years of age, plus or minus two months, and in the fifth grade. These two ages were selected in order to investigate whether differences in socioeconomic level effect differences in the pattern of test responses as compared with the intercorrelations of the subtests with those at seven and ten as published by Wechsler [9].

To determine socioeconomic level, the Warner-Meeker-Eells' Index of Status Characteristics was used [8]. This index uses four factors: occupation of father, source of income, house type, and characteristics of the dwelling area. The factors are rated and a weighted total of the ratings is obtained. The total is converted to social-class equivalence as upper-upper, lower-upper, upper-middle, lower-middle, upper-lower, lower-lower. Table 1 shows the social class ratings of the subjects in the study. There is no overlapping of the two groups. The subjects were selected only on the bases of sex, age, grade, and socioeconomic level.

Table 1
Distribution by Social Class Level

Status Index	Second grade		Fifth grade		Total
	Boys	Girls	Boys	Girls	
Upper Group					
Upper-Upper	0	0	0	0	0
Lower-Upper	1	0	2	2	5
Upper-Middle	6	9	5	8	28
Lower-Middle	3	1	3	0	7
Lower Group					
Upper-Lower	4	0	1	3	8
Lower-Lower	6	10	9	7	32
Total	20	20	20	20	80

The WISC was administered to all sub-

Table 2
Means, *SD*'s, Verbal and Performance Scaled Scores on the WISC

	Second grade			Fifth grade			Total group		
	M	*SD*	*t*	M	*SD*	*t*	M	*SD*	*t*
Full Scale									
Upper	114.4	9.49		115.65	11.14		115.02	10.37	
			4.20**			1.96			4.09**
Lower	101.8	8.99		108.15	12.45		104.98	11.31	
Verbal Scores									
Upper	72.65	8.67		78.0	9.28				
			3.94**			2.06			
Lower	61.65	8.98		71.0	12.06				
Performance Scores									
Upper	70.0	11.95		68.0	11.68				
			2.62*			1.63			
Lower	61.4	8.53		62.0	11.58				

**Significant at the 1 per cent level.
* Significant between the 1 and 5 per cent levels.

jects. The full scale of twelve subtests was used.

The second problem, that of "validity," was approached by giving all subjects the Metropolitan Achievement Test. The second-graders took Primary I Battery, Form R. The fifth-graders took the abbreviated form of the Intermediate Battery, Form R, which included tests of reading, vocabulary, arithmetic fundamentals, arithmetic problems, English, and spelling.

The third problem was investigated by means of intercorrelations of subtests of the WISC computed for the subjects in this study.

Results

I. Does Socioeconomic Level Effect Differences in Distribution of IQ's?

1. Analysis of variance using the method of repeated measurements on the same subjects, analogous to the method described by Edwards [1], shows that scores on the WISC varied with socioeconomic level, with sex, and with verbal and performance halves of the WISC.[1] Scores were analyzed in terms of sex, although, since the numbers were equal, this variable was not of primary concern in this study (Table 5, deposited with ADI).

2. Table 2 shows a difference in mean IQ for the total group, i.e., both grades combined. The mean of the upper economic group was superior to the mean of the lower economic group. The variability of the two groups as expressed by the *SD*'s did not differ significantly.

When the total group is divided by grade (or age), it is found that in the second grade

Table 3
Mean IQ and *SD* by Grade, Economic Level, Sex

	M	*SD*	*t*
Second Grade			
Upper			
Male	116.7	10.23	
			1.06
Female	112.1	9.04	
Lower			
Male	106.4	10.09	
			2.53*
Female	97.2	4.19	
Fifth Grade			
Upper			
Male	117.0	7.13	
			.51
Female	114.3	13.93	
Lower			
Male	111.0	11.85	
			.998
Female	105.3	12.38	

*Significant at better than the five per cent level.

[1]To reduce printing costs Tables 5, 6, 7, 8 and 9 have been deposited with the American Documentation Institute. Order Document 3837 from American Documentation Institute, c/o Library of Congress, Washington 25, D.C., remitting $1.00 for microfilm (images 1 inch high on standard 35 mm. motion picture film) or $1.00 for photocopies (6 x 8 inches) readable withhout optical aid.

(Table 2) the difference in means is again very significant, better than the 1 per cent level. There is only one subtest in which the lower group exceeds the upper group—Coding.

In the fifth grade, however, the differences in means are not significant although there is only one subtest in which the lower group exceeds the upper group—Digit Span.

Table 2 also shows the comparison of the verbal and performance scores on the WISC. The differences were significant in both verbal and performance scores in the second grade but in the fifth grade only the verbal IQ means differed.

3. Table 3 presents the differences in mean IQ for the two grades and two economic groups by sex. The difference is significant only in the lower economic group in the second grade.

Thus, in answer to the first question, differences in socioeconomic level effect differences in scores on the WISC for the total group studied. When the group is separated by grade (or age), however, these differences exist for the second grade (seven-year-olds) but are not significant for the fifth grade (ten-year-olds).

II. Does Socioeconomic Level Effect Differences in Prediction Value of the WISC?

1. The second major finding was that there is not a significant difference between the two economic groups as to predictive power of the WISC. Using Fisher's z transformation it is found that the difference between the correlations with achievement in the second grade ($r = .54$ and $r = .33$) is not significant at the 5 per cent level; nor is the difference ($r = .77$ and $r = .46$) significant for the fifth grade. Thus if achievement is used as the criterion, the WISC did not predict better for one economic group than the other.

III. Does Socioeconomic Level Effect Differences in the Pattern of Test Response as Indicated by Subtests and Intercorrelations between the Subtests?

This problem was divided into three parts:

1. In terms of intercorrelations, how does the sample used in this study compare with Wechsler's standardization data?

2. In terms of intercorrelations, how does this sample when separated by socioeconomic level compare with Wechsler's standardization data?

3. Are there differences in the intercorrelations of subtests between upper economic status and lower economic status?

1. *Comparison of whole group with standardization data.* The intercorrelations of the seven and one-half year old group were compared with those presented by Wechsler [9] and the intercorrelations for the ten and one-half year old groups with those presented by Wechsler [9]. (See Tables 6, 7, deposited with ADI.) The level of 5 per cent was computed by Fisher's z transformation as given by Edwards [1].

Those which differ are Arithmetic and Information, Similarities and Comprehension, Arithmetic and Vocabulary, Arithmetic and Digit Span, Arithmetic and Coding, Vocabulary and Digit Span, and Picture Arrangement and Object Assembly. Out of a total of sixty-six intercorrelations, nine in the seven and one-half year group differ at the 5 per cent level from Wechsler's.

At the ten and one-half year old level, the r's which differ from Wechsler's are Information and Arithmetic, Comprehension and Digit Span, Arithmetic and Vocabulary, Arithmetic and Digit Span, Arithmetic and Picture Arrangement, and Arithmetic and Coding. That is, six intercorrelations out of sixty-six differ at the 5 per cent level.

The striking observation in those tests which differ is that it is Arithmetic which so frequently enters into the intercorrelations. At seven years old, six out of the nine which differ are combinations including Arithmetic, and at ten years old, five out of the six include Arithmetic. In each of these cases, the intercorrelations in this study are lower than those presented by Wechsler. Since Arithmetic is so strongly influenced by school training, this may be a reflection of differences in educational systems.

Since only nine at the seven and one-half year level (out of sixty-six) and six at the ten and one-half year level (out of sixty-six) differ from Wechsler's data, we assume that our groups conform to the standardization data and that they are representative.

2. *Comparison of group separated on basis*

of socioeconomic level with standardization data. Each age group was separated according to socioeconomic level and the intercorrelations compared with Wechsler's standardization data for each age. Again no essential differences from the standardization group were found. (See Tables 8, 9, deposited with ADI.)

(*a*) In the upper economic group at seven and one-half years, those *r*'s which differ beyond the 5 per cent level are Information and Picture Arrangement, Information and Block Design, Information and Object Assembly, Information and Mazes, Arithmetic and Vocabulary, Arithmetic and Picture Completion, Arithmetic and Picture Arrangement, Arithmetic and Block Design, Similarities and Picture Arrangement, Vocabulary and Digit Span, Vocabulary and Object Assembly, Vocabulary and Coding, and Vocabulary and Mazes.

Those which differ in the lower group at seven and one-half years are Information and Arithmetic, Arithmetic and Vocabulary, Arithmetic and Digit Span, Arithmetic and Coding, Vocabulary and Digit Span, Digit Span and Object Assembly, Picture Arrangement and Object Assembly, and Object Assembly and Mazes.

Thus in the upper group thirteen intercorrelations out of sixty-six differ from Wechsler's; in the lower group, eight differ. Since the number of cases was so much smaller in this study than in Wechsler's group, it is assumed that there are no essential differences.

(*b*) At ten and one-half years, in the upper economic group, differences from Wechsler's are obtained in eight intercorrelations: Information and Arithmetic, Comprehension and Digit Span, Comprehension and Picture Completion, Arithmetic and Digit Span, Arithmetic and Picture Arrangement, Arithmetic and Coding, Similarities and Picture Completion, and Vocabulary and Block Design.

In the lower economic group, there are no *r*'s which differ significantly from the standardization data.

3. *Intercorrelations of subtests according to socioeconomic level.* The third comparison (Table 8, ADI) is of intercorrelations according to socioeconomic level. At seven and one-half years five *r*'s differ according to socioeconomic level: Information and Picture Arrangement, Arithmetic and Picture Arrangement, Vocabulary and Coding, Digit Span and Block Design, and Digit Span and Object Assembly.

At ten and one-half years, (Table 9, ADI) only two *r*'s differ according to socioeconomic level: Arithmetic and Digit Span, and Picture Arrangement and Object Assembly.

Table 4 shows the differences between economic groups in the subtests in standard score units. In general, it is the verbal items that show the greatest differences—Vocabulary, Similarities, Comprehension.

The results of this study, therefore, suggest that socioeconomic level effects differences (*a*) in magnitude of scores on the WISC but not (*b*) in pattern of scores and (*c*) that socioeconomic level does not effect differences in prediction value of the WISC.

Table 4

Scaled Score Differences between Economic Groups in Subtests

Subtests	Differences in economic groups in points	
	Second grade	Fifth grade
Information	49	18
Comprehension	43	36
Arithmetic	12	3
Similarities	58	43
Vocabulary	53	50
Digit Span	5	–10†
Picture Completion	10	12
Picture Arrangement	51	32
Block Design	48	23
Object Assembly	37	32
Coding	–9†	16
Mazes	53	8

†No sign indicates that the upper economic group had more points. A minus sign indicates that the lower economic group was superior.

Discussion

Several explanations might be suggested for the decrease in socioeconomic effects on Wechsler IQ with increase in age or grade. Does a selective factor operate so that only the brightest of the lower economic group persist to reach the same grade at the same age as the upper economic group? The duller ones may still be in the third and fourth grades and only the most capable regularly promoted.

Is it possible that the public schools are narrowing the differences occasioned by economic status in the home so that with increase in age, there is a decrease in the discrepancy between measured IQ in the economic groups? Many studies have shown a decrease in IQ with increase in age [3, 7]. However, the pupils in these studies were continuing in an extremely impoverished environment with a minimum of stimulation. The subjects in this study were in good public schools where there was constant encouragement and stimulation. This is an attractive explanation if its validity could be established.

A third explanation may be consonant with research at the University of Chicago [2, 4, 5, 6], that test items are constructed for and favor the middle and upper economic groups and penalize the lower group. Thus at an early age when the home environment is most influential, the lower group does poorly. At a later age, the middle-class influences of school, Sunday School, movies, etc., are beginning to be evidenced and may affect responses on the WISC.

Since this study suggests that socioeconomic differences as reflected in the WISC decrease with age, the WISC may become a preferred test in clinical and educational situations at older ages. Future studies might investigate this hypothesis at older ages than used in the present study.

Summary and Conclusions

Two groups of subjects, of upper and of lower economic level, 40 in each group, were selected on the bases of age, sex, grade, and socioeconomic status. One-half were seven and one-half years, plus or minus two months, in the second grade; one-half were ten and one-half years, plus or minus two months, in the fifth grade. One-half were boys; one-half were girls.

Socioeconomic level was determined by the Warner-Meeker-Eells' Index of Status Characteristics. There was no overlapping between the two groups.

The full scale of twelve subtests of the Wechsler Intelligence Scale for Children was administered. The Metropolitan Achievement Test was given to each subject.

The results were analyzed as a preliminary exploration toward determining whether socioeconomic level affects responses to the items of the WISC so that test differences between upper and lower socioeconomic levels are manifest in:

1. The distributions of IQ's.
2. The predictability of the WISC.
3. The pattern of test response as indicated by intercorrelations of the subtests (*a*) with Wechsler's standardization data and (*b*) between economic levels.

Scores on the WISC vary with socioeconomic level in respect to magnitude of score; the upper economic group was superior to the lower economic group. When the groups were divided by age or grade, the difference was significant at the seven and one-half year level, not significant at the ten and one-half year level.

The WISC did not predict better for one economic group than for the other.

Patterns of scores on the WISC did not appear to be differentiated by socioeconomic level.

Received July 21, 1952.

References

1. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
2. Eells, K., Davis, A., Havighurst, R. J., Herrick, V. E., & Tyler, R. *Intelligence and cultural differences.* Chicago: Univer. of Chicago Press, 1951.
3. Gordon, H. *Mental and scholastic tests among retarded children.* London: Board of Education, Education Pamphlet No. 44, 1923.
4. Haggard, E. A. Influence of culture background on test performance. Paper presented before Invitational Conference on Testing Problems sponsored by Educational Testing Service. New York, October 29, 1949.
5. Mitchell, J. V. Social class differences in mental test performance. Unpublished paper read at Midwest. Psychol. Ass., 1951.
6. Neff, W. S. Socio-economic status and intelligence. *Psychol. Bull.*, 1938, **35**, 727-757.
7. Sherman, M., & Key, C. B. The intelligence of isolated mountain children. *Child Develpm.*, 1932, **3**, 279-290.
8. Warner, W. L., Meeker, Marchia, & Eels, K. *Social class in America.* Chicago: Science Research Associates, 1949.
9. Wechsler, D. *Wechsler Intelligence Scale for Children.* New York: Psychological Corp., 1949.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

A Comparison of the Wechsler-Bellevue and ACE Tests on a University Counseling Center Group

Reed M. Merrill and Louise B. Heathers
University of Washington Counseling Center

As indicated by the many studies reviewed by Rabin [6], Watson [9], and Rabin and Guertin [7], the Wechsler-Bellevue Intelligence Scale (W-B) is very widely used as a measure of general ability. Relatively few studies, however, have been made regarding its validity for predicting academic performance at the college level. In this study W-B IQ's are compared with scores on the American Council on Education Psychological Examination (ACE). The ACE was selected because of its widespread use as the measure of scholastic aptitude given entering college freshmen. Two points are of interest: (*a*) the relative capacity of the two tests to predict academic achievement at the college level and (*b*) the relationship between the scores obtained on the two tests.

These relationships have been investigated in three previous studies. Anderson [1] gave the W-B and the ACE to all but two members of an entering freshman class in a woman's college. The 1940 ACE was given on entrance; the W-B, some time during the student's second semester. ACE raw scores were used in the correlations. The mean Verbal, Performance, and Full Scale W-B IQ's for this group of 112 *S*s were 116.0, 115.8, and 118.5, respectively. Sartain [8] gave the W-B and the 1942 ACE to 50 college freshmen. The sex of the *S*s was not specified. Presumably they were volunteers. ACE raw scores were used in the computations. The mean Verbal, Performance, and Full Scale W-B IQ's for this group were 115.4, 115.1, and 117.5, respectively. The measure of scholastic achievement used in both studies was the grade-point average (GPA) obtained by a student during his freshman year. Gerboth [3], in studying the relationship between Forms I and II of the W-B, gave Form I to 25 men and 25 women at some time during their college career. The median class placement of the *S*s was at the sophomore level but all classes were represented. Presumably these *S*s were volunteers. Some of these *S*s had taken some form of the ACE on entrance. The number of *S*s in the Form I group for whom ACE scores were available and the average time interval between the two testings are not known. Since various forms of the ACE had been taken by the *S*s, ACE centiles were used in the computations. The measure of scholastic achievement used was the GPA obtained during one semester, probably the semester during which the W-B was given. The mean Verbal, Performance, and Full Scale W-B IQ's for this group were 124.4, 122.1, and 125.6, respectively.

The study reported here was based on 251 male students who were part of a larger sample seen over a two-year period at a university counseling center. The total group of *S*s is described in detail elsewhere [5]. The 251 *S*s were those members of the original group who had taken some form of the ACE as entering freshmen at the University of Washington. Since the W-B was taken at the time the student came to the counseling center, the class level of the *S*s and the time interval between the two test administrations varied. When they were given the W-B, 55% of the *S*s were freshmen, 31.5% were sophomores, and 13.5% were upper-division students. Their median age at this time was 20.96 years with a *Q* of 2.26 years. The median time interval between the two tests was 11.02 months with a *Q* of 8.52 months. The measure of scholastic

performance was the GPA achieved in academic subjects for the freshman year. The mean GPA[1] was 2.10 with a standard deviation of .63 and a range from 0.0 to 3.8; the mean GPA for all freshmen at the University of Washington during the period covered by the study was 2.2.[2] Table 1 presents the mean W-B IQ's and mean local entering-freshmen centiles obtained by the group. Local centiles were used in all computations since various editions of the ACE had been taken by the *S*s and these centiles were more stable from year to year for the local population than the national centiles.

Table 1

Means and Standard Deviations on the W-B and the ACE Tests of Student Client Group (N=248)

	Wechsler-Bellevue IQ's		
	Verbal Scale	Performance Scale	Full Scale
Mean	117.70	118.21	120.07
σ	8.55	9.72	8.64
	Local ACE Centiles		
	Q Scale	*L* Scale	*T* Scale
Mean	52.25	47.75	49.05
σ	28.10	29.30	29.00

As may be seen in Table 2 the correlations found between GPA and both the W-B IQ's and the ACE scores tend to be quite similar in the four studies. In general the W-B Full Scale predicts college achievement as well as does the Total ACE. The Verbal Scale predicts achievement better than any of the other scales while the Performance Scale predicts less well than any of the other scales. The results on the ACE are in agreement with those reported by Berdie *et al.* [2]; ie., (*a*) the Linguistic (*L*) Scale predicts general achievement as well as the Total ACE Scale and somewhat better than the Quantitative (*Q*) Scale, and (*b*) the correlations between grades and ACE scores are consistently higher for the female than for the male groups. It is of interest that this sex difference does not appear in the W-B correlations with GPA.

[1]Since three students withdrew prior to obtaining any grades, all data involving GPA are based on 248 instead of 251 *S*'s.

[2]A grade of *C* has a point value of 2.0 at the University of Washington.

Table 2

Comparison of Correlations of W-B IQ's and of ACE Scores with GPA in the Four Studies

Study	GPA, W-B IQ			GPA, ACE Score		
	V	P	FS	Q	L	T
Anderson [1]	.52	.23	.45	.36	.48	.49
Student Clients*	.58	.17	.46	.27	.39	.40
Sartain [8]	.58	.35	.53	.35	.44	.43
Gerboth [3]	—	—	.29	—	—	.41

*All *r*'s significant above the 1% level.

Table 2 gives the significance of the differences between the correlations with GPA obtained in this study. In obtaining the *t* ratios, *r*'s were transformed into Fisher's *z*'s, a statistic whose sampling distribution is normal and independent of the magnitude of the universe *r*. Since the correlations compared were all based on the same group, McNemar's [4] procedure for estimating the standard error of the difference between *z*'s was employed. This technique takes into account the intercorrelations between the measures.

Table 3

Significance of the Differences between Correlations with GPA on the Student Client Group (N=248)

Comparison	$z_1 - z_2$	$\sigma_{z_1-z_2}$	*t*
FS vs. T	+.073	.061	1.20
V vs. FS	+.166	.045	3.69*
V vs. T	+.239	.062	3.85*
V vs. L	+.251	.067	3.75*
V vs. Q	+.386	.069	5.59*
V vs. P	+.491	.071	6.92*
P vs. FS	−.325	.041	7.97*
P vs. T	−.252	.070	3.60*
P vs. L	−.240	.075	3.20*
P vs. Q	−.105	.070	1.50
L vs. Q	+.135	.067	2.01
T vs. Q	+.147	.044	3.34*

*Significant above the 1% level.

The results suggest that the time required to give the Verbal Scale is time well spent. This scale tends to predict college achievement better than does the ACE and these predictions do not appear to be affected by sex differences. In addition the scale provides the counselor with the opportunity to observe the

quality of a student's behavior. Such observations are often as useful in counseling as are quantitative scores. One can speculate that academic predictions based on the Verbal Scale are best because it reflects, like GPA, both personality and ability factors.

The results also emphasize the need to interpret test scores in a very conservative manner and the need to utilize data from more than one source in making predictions about general college achievement. For the highest correlation reported, .58 between the Verbal Scale and GPA, the coefficient of alienation is .81. Predictions based on a correlation of this magnitude are only 19% better than chance. Further, it should be noted that most correlations obtained between college GPA's and tests of ability have even less predictive power, particularly when predictions are made beyond the freshman level.

Table 4 presents the intercorrelations between W-B and ACE scores found in the four studies. Here there is more variability in the results, particularly with those involving the Performance Scale. The most noteworthy finding is that, in all four studies, the Verbal Scale predicts the various ACE scores as adequately as does the W-B Full Scale. In most of the comparisons the Performance Scale predicts ACE performance less well than does the Verbal Scale. Table 5 shows the significance of the differences between the intercorrelations between W-B and ACE scores found in this study.

Table 4

Comparisons of Intercorrelations of the W-B IQ's and ACE Scores in the Four Studies

Correlations	Anderson [1] (N=112)	Student Clients* (N=251)	Sartain [8] (N=50)	Gerboth [3] (N=?)
V,Q	.49	.50	.50	—
V,L	.51	.54	.72	—
V,T	.54	.61	.68	.57
V,P	.34	.43	.53	—
V,FS	—	.81	.97	—
P,Q	.31	.42	.50	—
P,L	.19	.35	.54	—
P,T	.25	.43	.55	—
P,FS	—	.80	.84	—
FS,Q	—	.55	.55	—
FS,L	—	.54	.71	—
FS,T	.48	.61	.69	.58
Q,L	.61	.49	.60	—
Q,T	—	.79	.79	—
L,T	—	.91	.94	—

*All r's significant above the 1% level.

Table 5

Significance of the Differences between Intercorrelations between W-B IQ's and Local ACE Centiles for the Student Client Group (N=251)

Comparison	$r_1 - r_2$	$\sigma_{r_1-r_2}$	t
V,Q vs. FS,Q	—.069	.045	1.53
V,L vs. FS,L	.00	—	—
V,T vs. FS,T	.00	—	—
V,Q vs. V,L	—.055	.071	0.77
V,Q vs. V,T	—.160	.048	3.33*
V,L vs. V,T	—.105	.039	2.69*
P,L vs. V,L	—.238	.071	3.35*
P,T vs. V,T	—.249	.075	3.32*
P,Q vs. V,Q	—.101	.073	1.38
P,Q vs. P,L	+.082	.068	1.21
V,L vs. Q,L	+.068	.082	0.83
V,P vs. Q,L	—.076	.081	0.93

*Significant above the 1% level.

These intercorrelations are too low to support the assumption that, in normal persons, performance level on various kinds of ability tests should be highly similar. They are too high to support the assumption occasionally implied in test interpretation that the different scales of the two tests measure discrete abilities.

Summary

Correlations between college freshmen GPA's and the various W-B and ACE scales are reported. The findings of Anderson, Sartain, and Gerboth are reviewed and compared with the results found on 248 male student counseling center clients. Similar comparisons are made relative to the intercorrelations between the W-B and the ACE scales.

The *L* Scale of the ACE predicts the general academic achievement of college freshmen as well as the Total ACE and better than the *Q* Scale. Correlations between ACE scores and grades are higher for women than for men; this trend is not apparent on the W-B. The Verbal Scale of the W-B appears to be the best single predictor of the general academic achievement of freshmen. It also predicts the Q, L, and Total Scales of the ACE as ade-

quately as does the Full Scale of the W-B and predicts the L Scale about as adequately as does the *Q* Scale. The Performance Scale is least related to both academic achievement and the various ACE scales. It is possible that the reduced correlation between the Verbal and Performance Scales is due not only to the restricted range of talent in the college samples but also to the fact that, at this level of ability, the two scales do not measure similar factors. The data definitely support the use of the W-B as a predictor of the academic achievement of freshmen. This test, especially its Verbal Scale, predicts as well as the ACE, a test specifically designed to predict the academic achievement of college freshmen.

However, all of the obtained correlations are of such a magnitude that predictions regarding an individual student's academic achievement in freshmen college courses are subject to considerable error. Though data such as found in this study are a valuable aid in student counseling, one does well to note the variance that is unaccounted for in such results. Thus we believe that test scores are only part of the data that must be considered by a qualified counselor in helping students make decisions related to college achievement.

Received June 11, 1952

References

1. Anderson, E. E., *et al.* A comparison of the Wechsler-Bellevue, Revised Stanford-Binet, and American Council of Education tests at the college level. *J. Psychol.*, 1942, **14**, 317-326.
2. Berdie, R., Dressel, P., & Kelso, P. Relative validity of the Q and L scores of the ACE Psychological Examination. *Educ. psychol. Measmt*, 1951, **11**, 803-812.
3. Gerboth, Renate. A study of the two forms of the Wechsler-Bellevue Intelligence Scale. *J. consult. Psychol.*, 1950, **14**, 365-370.
4. McNemar, Q. *Psychological statistics.* New York: Wiley, 1949.
5. Merrill, R. M., & Heathers, Louise B. Centile scores for the Wechsler-Bellevue Intelligence Scale on a university counseling center group. *J. consult. Psychol.*, 1952, **16**, 406-409.
6. Rabin, A. I. The use of the Wechsler-Bellevue Scales with normal and abnormal persons. *Psychol. Bull.*, 1945, **42**, 410-422.
7. Rabin, A. I., & Guertin, W. H. Research with the Wechsler-Bellevue test: 1945-1950. *Psychol. Bull.*, 1951, **48**, 211-248.
8. Sartain, A. Q. A comparison of the new Revised Stanford-Binet, the Bellevue Scale, and certain group tests of intelligence. *J. soc. Psychol.*, 1946, **23**, 237-239.
9. Watson, R. I. The use of the Wechsler-Bellevue Scales: a supplement. *Psychol. Bull.*, 1946, **43**, 61-68.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

The Effect of Variations in Ambiguity on Projection in the Children's Apperception Test[1,2]

Edith A. Weisskopf-Joelson and David B. Lynn
Purdue University

How the degree of ambiguity of projective pictures affects the amount of projective material elicited by these pictures is a question which concerns the clinical practitioner as well as the theoretician. The former might be helped in the selection of effective pictures for projective techniques if he had the answer to this question. The latter might profit from gaining insight into the relationship between two important variables, namely, projection and ambiguity of the stimulus.

It is generally assumed that projection increases with increasing ambiguity. Experimental studies appear to corroborate this impression [1, 3, 4, 5, 6, 8, 10, 11, 15] with few exceptions [7, 13, 14, 16]. One of the exceptions is a study by Weisskopf [16], in which the ambiguity of TAT pictures was increased by two methods: (*a*) by incomplete tracing of the contours of the pictures, (*b*) by short exposure of the pictures. In both cases it was found that the more ambiguous pictures elicited less projection than the less ambiguous ones, when college students are used as subjects.

The present study attempts to test the validity of the above results for a different group of subjects, namely, for children.

Subjects and Materials

The subjects were 50 children, ranging in age from nine years to nine years and eleven months. All subjects attended the fourth grade of the same public school. Their IQ's on the Revised Stanford Binet Intelligence Scale, Form L, ranged from 83 to 154, with a mean of 116 and a standard deviation of 15.57.

Ten pictures were selected from Bellak's Children's Apperception Test (CAT) [2]. These pictures were reproduced by incomplete tracing of the contours, i. e., part of the contours were omitted in order to increase the ambiguity. Figure 1 shows Picture 6 of the CAT incompletely traced. The pictures used for control purposes were modifications of the same ten pictures of the CAT. They were ob-

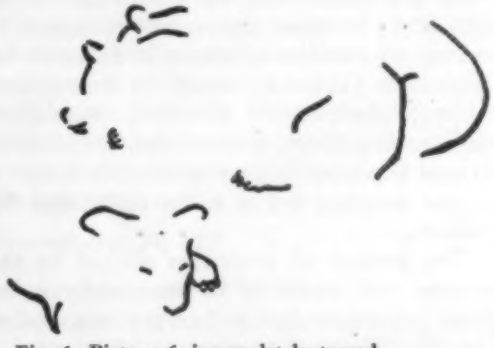

Fig. 1. Picture 6, incompletely traced.

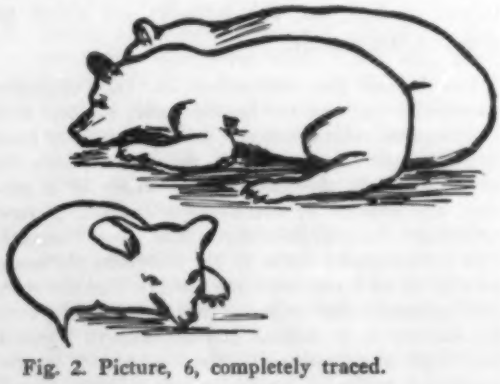

Fig. 2. Picture, 6, completely traced.

[1]The paper is based on a thesis presented to the Graduate School of Purdue University, Lafayette, Indiana by the junior author in partial fulfillment of the requirements for an M.S. degree in psychology. The study was directed by the senior author.

[2]This study was reported at the annual meeting of the Midwestern Psychological Association, Detroit, 1950.

tained by completely tracing the contours of the pictures, as illustrated in Figure 2. Thus, the material consisted of an experimental set of ten incompletely traced pictures and a control set of the corresponding ten completely traced pictures.

Procedure

The pictures were administered individually to each subject. The experiment was conducted in two sessions with a two-week interval between sessions. The incomplete pictures were administered in the first session and the complete ones in the second session. The order was never reversed because it can be expected that previous exposure of the complete pictures might decrease the ambiguity of the incomplete ones. However, it is reasonable to assume that the effect of repetition did not affect the results, since a previous unpublished study by Weisskopf shows that mere repetition of administration does not alter the transcendence index of incomplete pictures.

The instructions requested that the children "tell all about the picture." These instructions were used rather than the ones suggested by Bellak [2] in order to keep the procedure as similar as possible to the one followed by Weisskopf [17], who used the instructions "describe the picture." However, when using children as subjects, it was found advisable to change this latter phrase to one which has a similar meaning but is easier understood by children.

The amount of projection elicited by the pictures was measured by transcendence indices, a measure discussed in previous studies [16, 17, 18].

The transcendence index of a picture is obtained as follows: the subjects are asked to describe the picture.

> Even when the instructions call for objective description and not for fantasy, most subjects find it difficult to remain entirely descriptive. They tend to add projective material to their descriptions. *In order to obtain the transcendence index of a picture, the number of comments about this picture which go beyond pure description were counted. The transcendence index of the picture is the mean number of such comments per subject.* Pictures with low transcendence indices make it relatively easy for subjects to be factual and impersonal. Pictures with high transcendence indices make such impersonal observation difficult and, as it were, "lure" the subjects away from the objective path prescribed by the instructions and "force" him to project [17, p. 379].

The deviations from objective description were scored in terms of various categories such as the following:

> *Intraception* refers to the ascribing of emotions, desires, thoughts, fantasies, etc., to figures in the picture. . . . *Temporal Transcendence* refers to the inclusion of events which occur prior to, or after the event shown on the picture. . . . *Relationship* refers to the characterization of figures as related to other figures in the same picture. The great majority of relationship scores refer to kinship, as when one figure is described as the father, mother, son, daughter, husband, wife, etc., of another figure. . . . *Content of speech* refers to verbal statements made by pictorial figures. . . . *Atmosphere* refers to the characterization of the whole picture or parts of the picture in terms of the emotional response which it elicits in the observer. . . . *Imperative* refers to comments to the effect that a pictorial figure "should," "must," "is supposed to," act, feel, or think in a certain manner [17, p. 382-383].

The interjudge reliability, and the test-retest reliability of transcendence scoring have been previously discussed [18].

The score-rescore reliability of the junior author was found to be .98 for the complete pictures as well as for the incomplete ones.

The junior author and two additional judges scored the entire story material. Thus, each judge obtained a transcendence index for each picture. The mean of these three indices was used as the final transcendence index of the picture. The estimated reliability of this mean was found to be .96 for the complete pictures and .73 for the incomplete pictures [12]. The interjudge reliabilities of three judges ranged from .86 to .94 for the complete pictures and from .44 to .90 for the incomplete pictures.

Results

The mean transcendence index for the incomplete pictures is .60, for the complete pictures, 1.48. The difference between the two means is statistically significant well beyond the 1 per cent level. An individual comparison of each incomplete picture with the corresponding complete picture yields a difference in the same direction for eight out of ten pictures. Seven of these eight differences are sig-

nificant beyond the 1 per cent level, while one is significant between the 1 and 2 per cent level.

Discussion

The results of this study are in agreement with those of the above-mentioned study by Weisskopf, which investigates the same problem, using adults as subjects. "It seems that the subjects, when confronted with an indefinite and highly equivocal stimulus, produce less fantasy material than if confronted with a more definite picture" [16, p. 412-3]. The results of both studies are in agreement with Sears's [14] findings pertaining to doll play.

The implications of the results for the development of pictorial projective techniques have been presented in the earlier study:

> Can the results of this . . . study help the originators of pictorial projective tests to select pictures of high diagnostic value? The author believes that the answer is in the affirmative. . . .
>
> Are highly ambiguous or are highly structured pictures of greater diagnostic value? By increasing the degree of ambiguity of a picture . . . [by incomplete tracing], the production of fantasy material going beyond perception, such as the ascribing of emotions to human figures, or the imaginative creation of past and future events, is decreased. . . . Such a decrease would be a loss for diagnostic purposes, because transcendent material of this kind can be used as a basis for clinical diagnosis. However, this loss is probably accompanied by a gain in another direction. Even though highly ambiguous pictures do not elicit as much material *beyond* perception, they may enable the clinician to make diagnostic use of perception itself, that is, of what and how the subject perceives. With relatively ambiguous, unstructured pictures interpretation of this latter type may be more fruitful than with relatively definite pictures. High ambiguity of pictures increases the relative influence of individual differences on perception, as the protocols of the present study show.
>
> Thus, by using such relatively structured material as the . . . pictures [of the Children's Apperception Test], the clinician loses diagnostic material of one kind and gains diagnostic material of another kind. He *loses* the possibility of basing his interpretation mainly on *what* the subject sees in the picture, since most nonpsychotic subjects show considerable agreement as to what is actually shown. But he *gains* diagnostic material consisting of *additions* to perception, such as the imaginative creation of past and future events, or the ascribing of emotions and thoughts to human figures. In the Rorschach, material of this kind is rarely produced. According to the results of this study this comparative lack of transcendence in the Rorschach may be due to the fact that the Rorschach blots are so much more ambiguous than the . . . pictures [of the Children's Apperception Test]. "Rorschach interpretation . . . is based much more on *what* and *how* the subject perceives than what he *adds* to perception" [17, p. 380]. It "is based on transcendence of sensation rather than transcendence of perception" [17, pp. 380–381].
>
> It seems that in whatever direction the ambiguity of projective stimuli is varied, diagnostic material of one kind is lost and of another is gained. . . . Thus, the present study claims a decrease with increased ambiguity for one kind of projection (transcendence) only. The study of the factors causing the negative relationship between transcendence and ambiguity may be a fruitful objective for further research [16, pp. 414-415].

Summary

The amount of projection elicited by pictures was studied as a function of the ambiguity of the pictures.

The subjects were 50 nine-year-old children. Ten pictures from the Children's Apperception Test by Bellak were traced, whereby parts of the contours were omitted to increase the ambiguity. The amount of projection elicited by these pictures was compared with the amount of projection elicited by fully traced, and thus less ambiguous pictures .The pictures with higher ambiguity elicited less projection than the less ambiguous pictures. This difference was shown to be significant well beyond the 1 per cent level.

Received June 25, 1952.

References

1. Asch, S. E., Block, H., & Hertzman, M. Studies in the principles of judgments and attitudes: I. Two basic principles of judgment. *J. Psychol.*, 1938, **5**, 219-251.
2. Bellak, L., & Bellak, Sonya S. *Children's Apperception Test.* New York: C.P.S. Co., 1949.
3. Bruner, J. S., & Goodman, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33-44.
4. Carter, L. F., & Schooler, K. Value, need, and other factors in perception. *Psychol. Rev.*, 1949, **56**, 200-207.
5. Coffin, T. E. Some conditions of suggestion and suggestibility: A study of certain attitudinal and situational factors influencing the process of suggestion. *Psychol. Monog.*, 1941, **53**, No. 4 (Whole No. 241).
6. Hertzman, M. Studies in the principles of judgments and attitudes: III. The functional

equivalence of two differently structured references. *J. soc. Psychol.*, 1940, **12**, 3-19.
7. Levine, R., Chein, I., & Murphy, G. The relation of the intensity of a need to the amount of perceptual distortion: a preliminary report. *J. Psychol.*, 1942, **13**, 283-293.
8. Luchins, A. S. Social influences in perception of complex drawings. *J. soc. Psychol.*, 1945, **21**, 257-273.
9. Luchins, A. S. The stimulus field in social psychology. *Psychol. Rev.*, 1950, **57**, 27-30.
10. McClelland, D. C., & Atkinson, J. W. The projective expression of needs: I. The effect of different intensities of the hunger drive on perception. *J. Psychol.*, 1948, **25**, 205-222.
11. McGregor, D. The major determinants of the prediction of social events. *J. abnorm. soc. Psychol.*, 1938, **33**, 179-204.
12. Mosier, C. I. On the reliability of a weighted composite. *Psychometrika*, 1943, **8**, 161-168.
13. Schonbar, Rosalea A. The interaction of observer pairs in judging visual extent and movement: the formation of social norms in "structured" situations. *Arch. Psychol., N. Y.*, 1945, No. 299.
14. Sears, R. R. Influence of methodological factors on doll play performance. *Child. Develpm.*, 1947, **18**, 190-197.
15. Sherif, M. A study of some social factors in perception. *Arch. Psychol., N. Y.*, 1935, No. 187.
16. Weisskopf, Edith A. An experimental study of the effect of brightness and ambiguity on projection in the Thematic Apperception Test. *J. Psychol.*, 1950, **29**, 407-416.
17. Weisskopf, Edith A. A transcendence index as a proposed measure in the TAT. *J. Psychol.*, 1950, **29**, 379-390.
18. Weisskopf, Edith A., & Dunlevy, G. P. Bodily similarity between subject and central figure in the TAT as an influence on projection. *J. abnorm. soc. Psychol.*, 1952, **47**, 441-445.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

Sex Sequence in the Draw-a-Person Test and its Relation to the MMPI Masculinity-Femininity Scale

Samuel Granick
Washington, D. C.

and Leon J. Smith
St. Petersburg, Florida

Subjects are expected to draw a figure of their own sex as their first response to the examiner's direction to "draw a person" [2, 5, 6]. Levy [5] reports that of 5000 adults examined, 87 per cent drew their own sex. A study of 80 college students, ranging in age from 18 to 58 years with a median of 26, showed similar results, 76 per cent of the subjects drawing their own sex as their initial response [3]. With reference to children between the ages of 5 and 12 years, Jolles [4] found that both the boys and the girls drew their own sex predominantly when taking the H-T-P Test. Data obtained from a group of Serbian refugee children [1], on the other hand, indicated that while boys showed a preference for drawing males, girls under 10 tended to draw their opposite sex predominantly. The girls of 11 and 12 years of age, however, showed a preference for their own sex. How adolescents tend to respond has not as yet, to the authors' knowledge, been reported in the literature.

What the drawing of the same or opposite sex as the first response to the Draw-a-Person test implies clinically for an individual and generally for various age and sex groups appears never to have been systematically explored. Machover [6] and Buck [2] emphasize that in drawing the human figure an individual is probably making some kind of portrait of himself. In Machover's words, "The process of drawing the human figure is for the subject, whether he realizes it or not, a problem not only in graphic skill, but one of projecting himself in all of the body meanings and attitudes that have come to be represented in his body image." She goes on to mention that empirically she has found "evidence of some degree of sexual inversion . . . in records of all individuals who drew the opposite sex first" [6, pp. 101, 102].

The study herein reported is not intended as a test of Machover's entire hypothesis. Indeed that would be exceedingly difficult to accomplish since defining "some degree of sexual inversion" might prove elusive. Intended, however, is an investigation of the incidence of self-sex drawings as the initial response to the Draw-A-Person test on the part of a group of adolescents and a comparison of this with results obtained on a standard masculinity-femininity scale.

Method

The subjects for the experiment consisted of 185 male and 386 female students in the general psychology course during the fall semester of 1950-51 at the Florida State University. Since almost all students at the university tend to enroll in this course during their undergraduate years, this sample is probably representative, in terms of intelligence and cultural background, of the more than 6000 students who were enrolled at the school during that period and who, in turn, were mainly from the various sections of the state of Florida. The ages of the subjects were not recorded, but since almost all the students were in their sophomore year, it is probably safe to assume that few were above 22 or below 17 years of age. The vast majority ranged in age

between 18 and 20 years.

A booklet of four pages was prepared, two of the pages to be used for the drawings and the other two containing the 60 questions or statements of the Minnesota Multiphasic Personality Inventory Masculinity-Femininity Scale to which answers of "true" or "false" were to be given. The tests were administered by the junior author to the subjects in 16 groups of approximately 40 students each. When each subject had a booklet, the group was instructed as follows. "On the front page you are to draw a person. I am not interested in your ability to draw, but rather in how you attempt to draw a person. Please do not turn the pages of the booklet until told to do so." When all subjects had completed the first drawing, they were further instructed as follows. "Now turn to the next page. On this sheet you are to draw a person of the opposite sex to the one you have just drawn. If you drew a male on the front page, draw a female here. If you drew a female, now draw a male." When this drawing had been accomplished, the following directions were given in the administration of the M-F Scale of the MMPI. "Please place the letter *M* for 'male' or the letter *F* for 'female' in accordance with your sex in the space provided at the top left-hand corner of the third page. Now circle *T* for 'true' and the letter *F* for 'false' to these statements as they apply to you." Invariably students questioned whether or not they should append their names to the booklets. They were told it was not necessary to do so unless they wished to. Mention was also made that their responses would in no way affect their grades in the course.

Results

Of the 185 males, 163 or 88.1 per cent drew their own sex first. Similarly, the majority of the female subjects drew their own sex first (252 or 65.3 per cent of 386), but the proportion was significantly lower (*CR*=6.7, giving a p which is less than .01). The *SE*'s of the percentages were relatively low in each case (2.40 for the males and 2.42 for the females), thus indicating that the occurrence of the self-sex drawings for both groups was beyond chance expectation. These results agree fairly closely with Holtzman's report [3] showing that 85 per cent of the males (N=40) and 67.5 per cent of the females (N=40) drew their own sex as their first responses to the Draw-a-Person test. Generally, therefore, the sex sequence of the human-figure drawings differentiated the sexes on a group basis.

Table 1

MMPI M-F Scale Scores for Subjects

Subjects	*N*	Mean	*SD*	*t*	*p*
Males	185	26.30	4.02	29.6	<.001
Females	386	37.26	4.50		

The results of the MMPI M-F Scale were found also to separate the two sex groups as shown in Table 1.

It should be noted, furthermore, that relatively little overlapping occurred between the distributions of the two sets of scores.

Table 2

Sex Sequence in the Draw-a-Person Test and MMPI M-F Scale Scores

Subjects	No. Own Sex First	M-F Mean	M-F *SD*	No. Opposite Sex First	M-F Mean	M-F *SD*	*t*
Males	163	26.30	4.22	22	26.22	5.10	.07
Females	252	37.64	4.16	134	36.67	4.90	.18

The question now arises as to what relationship, if any, may exist between the incidence of self-sex drawings as a first response to the Draw-a-Person test and the ratings achieved on the M-F Scale. Calculation of the point biserial correlations for both sexes indicated that this relationship is probably of a zero order (r_{pbi}=.009 for males and r_{pbi}=.105 for females). A comparison of the M-F Scale means of the subjects of each sex who drew the same sex first with the means of those who drew the opposite sex first, as shown in Table 2, further substantiated this. In addition, the differences between the mean M-F Scale scores of those drawing either of the sexes first were in the direction of the sex of the subjects rather than of the sex drawn first. Thus a comparison of the scores of the males and females drawing the male figure first yielded a t of 19.9, showing a highly significant difference between the M-F Scale means. Similarly a t of 10.2 was obtained from the comparison of the M-F Scale scores of those subjects in both sex groups who drew the female figure first. Finally, a t of 8.9 was found between the M-F Scale scores of both sexes for those drawing the opposite sex figure first.

Discussion

Evident from the data is the fact that both the Draw-a-Person test and the MMPI M-F Scale serve to differentiate the male and female groups quite definitively. Apparently the tendency to draw the same sex first; however, is far stronger in men than in women. In these findings the present study is in essential agreement with results reported by other investigators. Added only is the fact that what is found among young children [4] and adults [3, 5] is also clearly the case with older adolescents attending college.

Not so clear, on the other hand, is the significance of the self- or opposite-sex drawings as the initial response to the Draw-a-Person test. Insofar as the MMPI M-F Scale is a measure of psychosexual identification, the data would tend to refute the hypothesis of Machover [6] and Buck [2] suggesting that sexual inversion may be involved in the choice of the opposite sex as the first response to the request of the examiner to "draw a person."

Normal cultural factors are probably among the important determinants of the trend noted in this and other reports for people to draw their own sex predominantly on the Draw-a-Person test. Those subjects who fail to follow the trend may, of course, be responding to different cultural influences, or to a variety of internal stimuli, which lead to the association of opposite sex as their initial reaction. Empirically Machover [6] and Buck [2] have found reason to believe that one's drawing of the human figure is essentially a self-portrait. But this may not be as generally true as they lead one to suppose. It is logical to suppose that the sex of the person one is inclined to draw as one's first response to the Draw-a-Person test may be related to internal and external stimuli far removed from psychosexual identification and self-portraiture. This is not to suggest, on the other hand, that in many instances conscious or unconscious portrayal of one's body image is not involved. Evident, however, is the need to investigate the fantasies and environmental factors affecting an individual at the time he is responding to the Draw-a-Person test.

Summary and Conclusions

The Draw-a-Person test and the MMPI M-F Scale were administered to a group of 185 male and 386 female college students taking the general psychology course. The sex of the figure drawn as the first response was noted in each instance and compared, in turn, with the scores made on the masculinity-femininity test. An analysis of the data yielded the following conclusions.

1. Male and female college undergraduates tended to draw their own sex predominantly when directed to "draw a person." The incidence of same-sex drawings was significantly greater among males than among females.

2. The MMPI M-F Scale significantly differentiated between the male and female subjects.

3. No relationship was found between the sex sequence of human figures drawn and scores on the masculinity-femininity measure.

Received June 4, 1952.

References

1. Baumgarten, F., & Tramer, M. Kinderzeichnungen in vergleichend psychologischer Beleuchtung. (Child drawings in the light of comparative psychological study.) *Z. Kinderpsychiat.*, 1943, **9**, 161-220. (Abstracted by: Florence L. Goodenough, & D. B. Harris. Studies in the psychology of children's drawings: II. 1928-1949. *Psychol. Bull.*, 1950, **47**, 369-433.)
2. Buck, J. N. The H-T-P Test. *J. clin. Psychol.*, 1948, **4**, 151-159.
3. Holtzman, W. H. The examiner as a variable in the Draw-a-Person test. *J. consult. Psychol.*, 1952, **16**, 145-148.
4. Jolles, I. A study of the validity of some hypotheses for the qualitative interpretation of the H-T-P for children of elementary school age: I. Sexual identification. *J. clin. Psychol.*, 1952, **8**, 113-118.
5. Levy, S. Figure drawing as a projective test. In L. E. Abt & L. Bellak, *Projective psychology.* New York: Knopf, 1950. Pp. 257-297.
6. Machover, Karen. *Personality projection in the drawing of the human figure.* Springfield, Ill.: Charles C Thomas, 1949.

Journal of Consulting Psychology
Vol. 17, No. 1, 1953

A Note on the "Tree and Trauma" Interpretation in the HTP[1]

Murray Levine and Eugene H. Galanter[2]
University of Pennsylvania and VA Mental Hygiene Clinic, Philadelphia

In his manual Buck [1, p. 390] suggests that blemishes drawn on the tree, such as scars, knotholes, broken branches, points of unusual shading or reinforcement, or weak or dead branches, may be representative of traumata which have occurred in the life of the person making the drawing. Furthermore he suggests [1, p. 340] that the ratio between the height of the traumatic indicator and the total height of the tree will yield the approximate age, plus or minus one year, of the occurrence of the trauma. The suggested interpretation has practical importance in terms of diagnosis, and it also contains implications for the theory of projection in expressive techniques. Specifically it would mean that the tree drawing has some relationship to the self concept as the self concept is projected into a "life space." It is then of considerable interest to treat the interpretative suggestion as an hypothesis and to attempt an objective verification of that hypothesis.

The clinical evidence related to this suggestion is seemingly derived from the *post hoc* examination of the individual's life history for incidents considered traumatic by the individual. The interrogation of the subject involves suggesting the age of traumatic occurrence to him and his response is therefore *not* independent of the indicator in the drawing. Furthermore the criteria for a traumatic occurrence are not specified and any unusual occurrence mentioned by the subject is accepted as a valid "trauma." An objective test of the validity of the interpretation requires that both the trauma and its date of occurrence be specified independently of the drawing.

The writers obtained drawings of a tree from 27 hospitalized, paraplegic veterans assuming that the severity of the injury and the circumstances under which it occurred, as well as the complete reorientation to life the injury requires, make it eligible as a trauma. In addition one can specify the exact date of occurrence of the trauma. Both the trauma and the time of occurrence are in this way specified independently of the drawing. The instructions were simply "Draw a tree." The population ranged in age from 18 to 40, the median age being 26 years and the duration of injury ranging from 6 months to 10 years, with a median duration of 15 months.

The presence of traumatic indicators on the tree was independently rated by the authors according to the general criteria stated by Buck. The raters agreed as to presence or absence of indicators in 82 per cent of the cases. The degree of agreement is significantly greater than chance agreement of 50 per cent. Of the 27 drawings rated, only 7 contained unequivocal traumatic indicators. The fact that only a small percentage of this obviously traumatized population drew indicators which could be reliably rated as such suggests that either the hypothesis has dubious validity, or that other unknown personality factors contribute to the appearance or nonappearance of such indicators on tree drawings. The age of occurrence of the trauma was calculated in each of these 7 drawings and compared with the actual age of traumata as obtained from the patient's records. The correspondence was poor. None of the cases fell within the two-year

[1]Reviewed in the VA and published with the approval of the Chief Medical Director. The statements and conclusions published by the authors are the result of their own study and do not necessarily reflect the opinion or policy of the VA.

[2]The authors wish to thank Dr. L. J. Hatterer of the Kingsbridge VA Hospital, New York, who was instrumental in helping to gather the reported data.

limit suggested by Buck. However, a rank-order correlation between the estimated age of occurrence obtained from the drawing and the actual age of occurrence obtained from hospital records proved to be plus .73; significant for the 7 cases at approximately the .05 level of confidence. In addition to the few cases, confidence in the size of this coefficient is lessened by the presence of split ranks. These results, however, suggest that the hypothesis may have some degree of validity in terms of a broad approximation of the time of injury, but statements concerning more exact localization in time are not supported by our data.

Since only a small percentage of this traumatized population spontaneously drew traumatic indicators, future research will require dealing with a very much larger parent population in order to obtain sufficient cases to test the hypothesis adequately. Verification of the interpretation, even in the form suggested by the present results, would seem to have interesting theoretical consequences for the general concept of projection in expressive techniques, and specifically for the notion that the self concept can be studied by this and similar methods. Others who may be working with similarly traumatized individuals might profitably repeat the study with larger samples.

Received June 6, 1952.

Reference

1. Buck, J. N. The H-T-P technique: A qualitative and quantitative scoring manual. *J. clin. Psychol.*, 1948,**4**, 317-396.

NEW BOOKS AND TESTS

Books

Ashby, W. Ross. *Design for a brain.* New York: John Wiley & Sons, 1952. Pp. ix + 260. $6.00.

This is another excellent addition to the list of distinguished treatises on cerebral mechanisms that Wiley has been publishing. The problem that Ashby has set himself is "to deduce from the observed facts" what are the necessary properties of a nervous system that acts "at once mechanistically and adaptively." The book is well written, the thesis clearly and tightly developed, and there is much more use of and integration with specific biological findings than one is apt to get in theoretical works of this sort. It remains, however, of necessity a difficult book which merits a larger audience than it will probably get. Its appeal will be greater to students of the logic of science and of system construction than it will be to practicing clinicians. The appendix contains a 50-page restatement of the thesis in mathematical form.—*W. A. H.*

Baumgarten, Franziska (Ed.) *La psychotechnique dans le monde moderne.* C. R. du IX^{e} Congres International de Psychotechnique. Paris: Presses Universitaires de France, 1952. Pp. xi + 630. 2,000 fr.

This is the report of the 9th International Congress of Psychotechnology held in Berne, Switzerland, in 1949, fifteen years after the last previous Congress. In addition to the welcoming addresses, and a report by the General Secretary of the preceding fifteen years of the work of the International Association of Psychotechnology, the volume includes over 100 papers presented by persons representing 23 different countries. More than half of the papers are in French, the others are in German, English, or Italian. These papers are presented under fourteen topical headings representing a wide array of problems and applications. Although containing but few detailed research reports, the volume offers an unusual opportunity to become acquainted with the current thinking and practice of our colleagues in other lands. Only seven of the papers were by Americans!—*E. L. K.*

Note.—The reviews were prepared by the Editor and the Associate Editors, who may be identified by their initials.

U. S. Federal Security Agency. Public Health Service. National Institute of Mental Health. *Mental health and motion pictures.* Publ. Hlth Serv. Publ. No. 218. Washington: U. S. Government Printing Office, 1952. Pp. 124. $.30.

For anyone who has the responsibility of publicizing or training in the various aspects of mental health this guide to 106 films is a positive must. The films are grouped under Child and Family, Marriage, Mental Health and Schools(for teachers, and for students), Mental Health Problems, Physiology of Human Reproduction, and Miscellaneous. The synopses clearly present the description of the film, its aims, and other information pertinent to its use. The format is excellent, giving running time, distributor, television information, followed by ample space for personal comments by the local viewer. The "Suggested Audiences" contains valuable information. Care has been taken not only to state the audience for which the film was planned but also, in the case of films whose presentation needs more careful controls, the type of control needed such as the presence of a trained professional worker. The book contains a list of agencies designated as State Mental Health Authorities, names and addresses of distributors, and an excellent index.——*B. M. L.*

Frosch, John. (Ed.) *The annual survey of psychoanalysis.* Vol. I, 1950. New York: International Universities Press, 1952. Pp. xvi + 556. $10.00.

"This is the psychoanalytic literature of 1950, all of it. . . . " With this opening statement, a new survey venture in psychoanalytic literature was launched. Apparently other annual surveys are to follow. The publication aims at abstracting all the periodical literature of the field and many of the books for a given year; it does not aim at critical reviews. Eleven of the twelve chapters deal systematically with articles in the field, while the last chapter abstracts 12 of the 42 books published during the year 1950. Some idea of the subject organization can be gained from several chapter titles: History, Methodology, Clinical Studies, Child Analysis, Psychoanalytic Therapy, Applied Psychoanalysis, etc. Many contributors are listed as abstractors, and most of the chapters have different authors. The job done in this survey is tremendous, but even more important is the quality of most of the abstracts. Comparing the abstracts with some of the complete articles creates great respect for the

compilers of this publication. Not infrequently is an article improved in readability and clarity by the abbreviation; writers in the psychoanalytic field are not as a rule adherents of the Law of Parsimony. The *Annual Survey of Psychoanalysis* is a welcome addition to psychological and psychiatric literature. Psychoanalysis will probably win more friends and more understanding through this volume than through the original material it attempts to abstract.—*M. K.*

Hirsh, Ira J. *The measurement of hearing.* New York: McGraw-Hill, 1952. Pp. ix + 364. $6.00.

The fields of experimental psychology and clinical audiometry are nicely integrated in this comprehensive and useful book on hearing. The introductory chapters give enough background on psychophysics, the physics of sound, and electroacoustic devices to serve the needs of readers who have at least a minimum acquaintance with the physiology and psychology of audition. The major problems taken up in the remaining chapters include the measurement of the hearing of pure tones and of speech, masking, fatigue, differential sensitivity, binaural hearing, bone conduction, and the use of conditioned response methods in audiometry. Because hearing loss is often related to a person's broader social adjustments, clinical psychologists have good reason to maintain more than a casual acquaintance with audiometry. They will find Hirsh's book an excellent resource.—*L. F. S.*

Kornitzer, Margaret. *Child adoption in the modern world.* New York: Philosophical Library, 1952. Pp. xiii + 403. $4.50.

The greater part of this book, written by the press officer of a group of British adoption societies, concerns present practices with detailed description of legal procedures in Great Britain. There is also some routine discussion of adoption problems generally, selection of adoptive parents, telling the child, the unadoptable child, and so on. Adoption first became legal in Great Britain in 1926; in 1949, seventeen thousand adoption orders were made. About one-third of the book discusses adoptive practices in other parts of the British Commonwealth, and brief reference is made to the situation in other countries. Two chapters are devoted to the United States. These summarize the legal situation in the various states and give some over-all comparison with that in Great Britain, but contain chiefly a "mosaic of brief quotations . . . because only in this way is it possible for a British writer to present, primarily for British readers and without causing offence to American ones anything like a valid picture of what goes on in the American adoption field today."—*A. R.*

Kuhlen, Raymond G., & Thompson, George G. (Eds.) *Psychological studies of human development.* New York: Appleton-Century-Crofts, 1952. Pp. xiii + 532. $3.50.

Currently, source books for introductory courses in psychology are finding their way into the classroom in increasing numbers. This paper-bound volume is a collection of 71 abridged and edited primary source materials in the general area of developmental psychology—here defined as extending from infancy through senescence. The subject matter covered includes physical factors in development, learning, intellectual development, social-cultural conditions, interest patterns, vocational orientation, and personal-emotional adjustment. The papers, originally published between 1928 and 1951, represent not only different research areas, but different methodologies, with the result that the studies selected will be considered of uneven quality by many readers. Although many studies now considered "classics" are included (e.g., Gesell, Jones & Conrad, Terman, Hartshorne & May, and Levy), it is inevitable that most readers will find some of their particular favorites missing. Although in general the abridgment and editing are done with care, it is questionable whether some of the more extensive abbreviations are wise (as, the reduction of Levy's book, *Maternal Overprotection,* to eight pages). These problems intrinsic to source books, however, should not detract importantly from the usefulness of this volume as a teaching aid in introductory courses in developmental psychology.—*A. M.*

Wolfle, Dael. (Chairman) *Improving undergraduate instruction in psychology.* New York: Macmillan Company, 1952. Pp. vii + 60. $1.25.

This is a 60-page report of a study group which attacked the problems of the objectives and the curriculum of undergraduate instruction in psychology. The members of the committee are well known and respected by their colleagues and the report will carry weight. A recommended curriculum with core, special interest, advanced, and integrative courses is presented for the typical liberal arts undergraduate. It is frankly a psychology- or content-centered curriculum rather than a student- or problem-oriented one. The educational objectives of personal growth and social adjustment are considered but rejected in favor of knowledge, habits of thought, and attitudes and values. The report states further that "it is probably desirable to avoid direct frontal efforts to modify student's habits and attitudes." Little space is given to indirect and student-centered means of dealing with attitudes and values that do not employ the methods the committee rightly banned such as "pressure," disregard for student rights, or reformation. Although a traditional and conservative educational climate pervades the report, there is an occasional hint that in committee discussion the realities of student motivation, individual differences, and present societal needs occasionally arose. The pub-

lished report, however, does not seem to give much significance to these influential factors which will determine how well their educational objectives can be met. This report can render a real service if it provokes debate and research on both the potential role of psychology in a liberal education and the means by which this can be realistically accomplished with our present student body. It will serve also as a target of attack for those who feel along with this reviewer that the psychologist has a special responsibility and competency adroitly to assist students to meet their personal and social needs, and who also maintain that this is a primary educational goal and basic to a liberal education.—*F. McK.*

Tests

Bennett, George K., Seashore, Harold G., & Wesman, Alexander G. *Differential Aptitude Tests.* Manual (2nd Ed.), pp. 77 ($2.00). New York: Psychological Corp., 1952.

The *Differential Aptitude Tests* have been provided with an entirely new manual which replaces the original loose-leaf edition of 1947 and its five supplements issued between 1948 and 1951. In addition to the inevitable directions for administration, scoring, and use, the new manual contains unusually comprehensive data about the tests. Norms are now based on groups of 2100 to 7400 for Form A, and of 350 to 1075 for Form B, for each sex and for each grade from 8 to 12, separately. The validity of the *DAT* is described in terms of its prediction of subsequent course grades, achievement test scores, and postschool careers. There are about 20 pages of tables of validity coefficients. Also presented with more than minimum adequacy are reliabilities, standard errors of measurement, and the intercorrelations of the tests. Even more detailed research data have been deposited with the American Documentation Institute so as to be available to technically trained users. The *Manual* can serve as model for other test constructors, and as an adjunctive text for courses in measurement.—*L. F. S.*

Books Received

Adler, Leta M., Coddington, James W., & Stewart, Donald D. *Mental illness in Washington County, Arkansas: Incidence, recovery, and posthospital adjustment.* Little Rock, Ark.: Univer. of Arkansas, 1952. Pp. 74.

Brennan, Robert E. *General psychology.* (Rev. Ed.) New York: Macmillan, 1952. Pp. xxxii + 524. $5.50.

Coombs, Clyde H. *A theory of psychological scaling.* Engng Res. Bull. No. 34. Ann Arbor, Mich.: Univer. of Michigan, 1952. Pp. vi + 94. $1.75.

Eisler, Robert. *Man into wolf, an anthropological interpretation of sadism, masochism, and lycanthropy.* New York: Philosophical Library, 1952. Pp. 286. $6.00.

Lindner, Robert. *Prescription for rebellion.* New York: Rinehart, 1952. Pp. 305. $2.75.

Maslow, Paul. *Powers of the mind.* Brooklyn, N. Y. (16 Court St.): Author, 1952. Pp. 288. $3.50.

Meerloo, Joost A. M. *Conversation and communication, a psychological inquiry into language and human relations.* New York: International Universities Press, 1952. Pp. x + 245. $4.00.

Palmer, Harold. *The philosophy of psychiatry.* New York: Philosophical Library, 1952. Pp. ix + 70. $2.75.

Powys, John Cowper. *The inmates.* New York: Philosophical Library, 1952. Pp. 318. $4.50.

Helping parents understand the exceptional child. Proc. Annu. Conf. Educ. Except. Child. Langhorne, Pa.: The Woods Schools, 1952. Pp. 42.

NEUROLOGY & PSYCHIATRY

Section VIII of Excerpta Medica

The Neurology Section includes abstracts of all articles on neurological surgery in addition to general clinical neuro-physiology.

The section on Psychiatry covers the topics usually contained within the meaning of the word. In addition, the following special branches are dealt with fully:

* **Psychoanalysis, individual and analytical psychology**
* **Sexology, criminology, alcoholism, and drug addiction as related to psychiatry**
* **Social and industrial psychology and psychiatry, vocational guidance and mental testing**
* **Psychometrics**
* **Personality testing and the Rorschach test**
* **Heredity and statistical studies as they apply to psychiatry**
* **Mental defect and epilepsy**

Annual volume of 1000 pages, including an index classified by author and subject.

Subscription price $22.50 per annual volume

EXCERPTA MEDICA SERVICE CORPORATION

280 Madison Ave. **New York 16, N.Y.**

Please enter my subscription to Section VIII, NEUROLOGY AND PSYCHIATRY, OF EXCERPTA MEDICA. I enclose $22.50.

Name.....

Address.....

.....

Distinctive McGRAW-HILL Books

PSYCHOANALYTIC THEORIES OF PERSONALITY

By GERALD S. BLUM, University of Michigan. *McGraw-Hill Publications in Psychology.* In press

This new text seeks to present descriptive yet concise accounts of expert psychoanalytical opinions on personality development, organized according to age levels. The exposition is factual and objective. In addition to the consideration of leading opinions, the author includes a section of "Notes" at the end of each chapter which offer brief resumés of existing experimental data, suggested evidence from related fields, inconsistencies, and comparisons of overlapping views.

THE PSYCHOLOGY OF INDUSTRIAL RELATIONS

By C. H. LAWSHE, Purdue University. In press

Presented in an unusual, non-technical, and practical manner, this book shows the non-psychologists and even those who have never studied psychology the things industrial psychology offers. Aimed specifically at supervisors and managers in industry, it also is useful to the college student studying management.

PRATICAL PSYCHOLOGY. New 2nd Edition

By KARL S. BERNHARDT, University of Toronto. 337 pages, $3.75

Designed for students who are doing work at a post secondary school level, this text provides a sound treatment of psychology, using a practical approach. Simple in treatment, with down-to-earth applications, *Practical Psychology* covers most of the usual topics in introductory psychology, but adds chapters on child training, mental hygiene and personal adjustment. A chapter on aptitude measurement and another on personal happiness and efficiency have been added.

THE MEASUREMENT OF HEARING

By IRA J. HIRSH, Washington University and Central Institute for the Deaf. *McGraw-Hill Publications in Psychology.* 364 pages, $6.00

A text for clinical otologists, experimental psychologists, acoustics or communications engineers, and others who do experimental work on hearing. It brings together basic, experimental information about acoustics, electroacoustic equipment, psychology of hearing, etc., and applies this information to the several facets of the measurement of hearing.

Send for copies on approval

McGRAW-HILL BOOK CO., INC.

330 West 42nd Street, New York 36, N.Y.